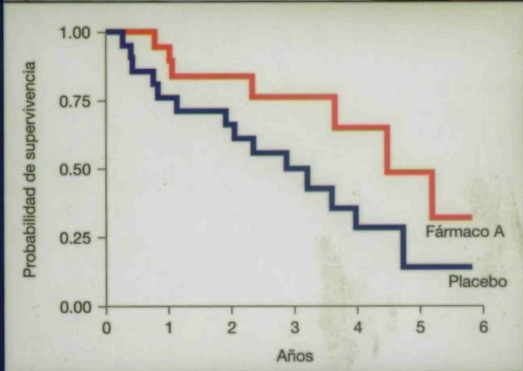




Agencia Internacional de Investigación sobre el Cáncer
Organización Mundial de la Salud

Epidemiología del Cáncer: Principios y Métodos



Isabel dos Santos Silva

Epidemiología del Cáncer: Principios y métodos

Agencia Internacional de Investigación sobre el Cáncer

La Agencia Internacional de Investigación sobre el Cáncer (IARC) fue establecida en 1965 por la Asamblea Mundial de la Salud, como una organización financiada independientemente dentro del ámbito de la Organización Mundial de la Salud. La sede de la Agencia está en Lyon, Francia.

La Agencia realiza un programa de investigación que se concentra especialmente en la epidemiología del cáncer y el estudio de carcinógenos potenciales en el ambiente humano. Sus estudios de campo se complementan por investigación biológica y química llevada a cabo en los laboratorios de la Agencia en Lyon, y por acuerdos de colaboración en investigación, en instituciones de investigación nacionales en muchos países. La Agencia también lleva a cabo un programa para la educación y formación de personal para la investigación sobre el cáncer.

Las publicaciones de la Agencia están destinadas a contribuir en la distribución de información sobre diferentes aspectos de la investigación sobre el cáncer. La información sobre las publicaciones de la IARC y cómo solicitarlas está disponible por Internet en: <http://www.iarc.fr/>



**Organización Mundial de la Salud
Agencia Internacional de Investigación sobre el Cáncer**

Epidemiología del cáncer: Principios y métodos

Por
Isabel dos Santos Silva

Agencia Internacional de Investigación sobre el Cáncer
Lyon, Francia
1999

Publicado por la Agencia Internacional de Investigación sobre el Cáncer,
150 cours Albert Thomas, 69372 Lyon cédex 08, Francia

© Agencia Internacional de Investigación sobre el Cáncer
Lyon, Francia. 1999

Distribuido por IARC Press (fax: +33 4 72 73 83 02; E-mail: press@iarc.fr)
y por Distribución y Ventas de la Organización Mundial de la Salud, CH-1211
Ginebra 27 (fax: +41 22 791 4857)

Las publicaciones de la Organización Mundial de la Salud disfrutan de protección de derechos de copia
acuerdo con las provisiones del Protocolo 2 en la Convención Universal de Derechos de Copia.

Todos los derechos son reservados.

Las denominaciones empleadas y la presentación del material en esta publicación no implican la expresión de una opinión por parte del Secretariado de la Organización Mundial de la Salud referente a la situación legal de un país, territorio, ciudad o área o de sus autoridades, o referente a la delimitación de sus fronteras o límites.

La mención de compañías específicas de productos de algunos fabricantes no implica que estén aprobados o recomendados por la Organización Mundial de la Salud con preferencia a otros de naturaleza similar que no se mencionan. A excepción de errores y omisiones, los nombres comerciales de productos se distinguen por letras iniciales en mayúscula.

Los autores son los únicos responsables de las opiniones expresadas en esta publicación.

La Agencia Internacional de Investigación sobre el Cáncer acepta solicitudes de permiso para reproducir o traducir sus publicaciones, parcial o totalmente. Las solicitudes y preguntas se deben dirigir al Servicio Editorial y de Publicaciones, Agencia Internacional de Investigación sobre el Cáncer, que le proporcionará la última información sobre cambios realizados en el texto, planes de nuevas ediciones y reimpressiones y traducciones ya disponibles.

Guía catalográfica de publicaciones

Epidemiología del Cáncer: Principios y métodos / Isabel dos Santos Silva
Versión española de la obra original: Cancer Epidemiology: Principles and Methods /
Isabel dos Santos Silva

Coordinación de la versión española: Dr. F. Xavier Bosch, Institut Català d'Oncologia
Traducción: Sofía Moreno Aguado

Revisión de textos: Dr. Xavier Castellsagué, Dr. Victor Moreno, Dra. Silvia de Sanjosé, Dra. Josepa Ribes
Dr. J. Alfons Espinás, Dra. M. Jesús Quintana, Pilar A. Martínez
E-mail: serc@ico.scs.es

1. Neoplasias – epidemiología

ISBN 92 832 0407 7 (Clasificación NLM W1)

Impreso por: Thau, sl., Arístides Maillol, 5, C1 - 08028 Barcelona (España)
D.L.: B-19.726-99

Indice

Prefacio	vii
Sobre este libro	ix
Nota de los traductores y Glosario	xi
1 Introducción a la epidemiología del cáncer	1
2 Determinación de exposiciones y resultados	11
3 Descripción y presentación de datos	47
4 Determinantes de la aparición de enfermedad y otros acontecimientos relacionados con la salud	61
5 Revisión de diseños de estudios	89
6 Evaluación del papel del azar	109
7 Estudios de intervención	141
8 Estudios de cohorte	173
9 Estudios caso-control	199
10 Estudios transversales	225
11 Estudios basados en datos de rutina	243
12 Introducción al análisis de supervivencia	279
13 Interpretación de estudios epidemiológicos	295
14 Tratando con la confusión en el análisis	325
15 Tamaño de un estudio	353
16 Prevención del cáncer	377
17 Papel de los registros del cáncer	409
18 Diseño, planificación y realización de la investigación epidemiológica	431
Referencias	455
Índice de materias	465

Prefacio

Dentro del programa de la IARC de cursos de formación, se otorga especial prioridad a los cursos de epidemiología del cáncer, especialmente en países en desarrollo. Los estudios epidemiológicos basados en la población son la única fuente fiable de la aparición y desenlace del cáncer y son el fundamento de cualquier programa nacional o regional de control del cáncer. Durante años, la Agencia Internacional de Investigación sobre el Cáncer se ha enfrentado al problema de seleccionar un libro de texto adecuado para sus cursos de formación. Para resolver esta dificultad, decidimos publicar un libro de texto adecuado tanto para el personal docente como para el lector no iniciado. Isabel dos Santos Silva, de la Escuela de Londres de Higiene y Medicina Tropical, una epidemióloga destacada con amplia experiencia docente, ha realizado esta labor de manera admirable, con una estrecha asesoría de expertos externos y un Comité Asesor de la IARC (véase más adelante).

Esperamos que este libro de texto guíe a los estudiantes y profesionales sanitarios que deseen entender los principios básicos y los métodos utilizados en la epidemiología del cáncer.

P. Kleihues
Director, IARC

Comité Asesor de la IARC

Nubia Muñoz
D. Maxwell Parkin
Jacques Estève
Paolo Boffetta

Sobre este libro

Este libro está dirigido principalmente a estudiantes de medicina y de salud pública, médicos, profesionales sanitarios y aquellos que quieren entender los principios y métodos utilizados en epidemiología del cáncer. Su objetivo no es convertir a los lectores en expertos en epidemiología, sino en hacerlos competentes en el empleo de herramientas epidemiológicas básicas y capaces de ejercer un juicio crítico al evaluar los resultados publicados por otros.

El libro está diseñado para proporcionar una comprensión fácil de los conceptos básicos y de los métodos mediante ejemplos ilustrativos, con frecuencia utilizando datos reales de investigación. Mi intención fue asegurar que el material fuera accesible a cualquier lector nuevo en el tema. El conocimiento de estadística es indispensable para la realización, análisis e interpretación correctos de los estudios epidemiológicos. Por tanto, también se presentan los conceptos (y fórmulas) estadísticos, pero se presta mayor interés a la interpretación de los datos que a los cálculos en sí.

El libro está dividido en 18 capítulos. Los seis primeros introducen los principios básicos de epidemiología y estadística. Los capítulos 7 – 13 tratan con mayor profundidad cada uno de los diseños de estudio y la interpretación de sus observaciones. Los capítulos 14 y 15 tratan conceptos estadísticos más complejos y se pueden omitir en una primera lectura. El capítulo 16 trata aspectos metodológicos en la prevención del cáncer, incluyendo la prevención primaria, la detección precoz (actividades de cribaje) y la prevención terciaria. El capítulo 17 revisa el papel de los registros del cáncer en la epidemiología y prevención del cáncer. Por último, el capítulo 18 comenta los aspectos logísticos y las consideraciones prácticas que se deben tener en cuenta en el diseño, planificación y realización de cualquier tipo de investigación epidemiológica.

Pocas, si alguna, de las ideas y conceptos en este libro son originales. Muchas de ellas proceden de mi propia experiencia docente en la Escuela de Higiene y Medicina Tropical de Londres y, especialmente, de mi función como organizadora del curso intensivo de epidemiología y estadística médica de la Escuela. Me siento en deuda con muchos compañeros y estudiantes con los que he trabajado los últimos años por todo lo que he aprendido de ellos. Sería imposible nombrar a todos los que, de una manera u otra, me ayudaron a escribir este libro – sólo puedo esperar que ellos sean conscientes de mi gratitud. No obstante, estoy especialmente agradecida por la ayuda de mis compañeros (y buenos amigos) Bianca De Stavola

y Esä Läärä por las horas que pasaron leyendo los primeros borradores y por sus útiles comentarios, sugerencias y discusiones estimulantes – ¡realmente aprendí mucho de ellos! También querría agradecer a la IARC por darme la oportunidad de escribir este libro y, en particular, a los miembros del comité editorial (Drs. Jacques Estève, Nubia Muñoz, Max Parkin y Paolo Boffetta), al Dr. Rengaswami Sankaranarayanan y al Dr. Martyn Plummer, por sus sugerencias y comentarios en las primeras etapas del proyecto. También agradezco por toda la ayuda y ánimo prestados por mis compañeros de la Escuela de Londres de Higiene y Medicina Tropical, especialmente por los comentarios atentos sobre versiones anteriores de varios capítulos proporcionados por el Catedrático Peter Smith, la Dra. Noreen Maconochie y el Dr. Punam Mangtani. También quiero agradecer a la Sra. Maria Quigley por su permiso para incluir uno de sus ejemplos docentes en el capítulo 14. Parte del material presentado en los capítulos 6 y 12 fue desarrollado inicialmente como material docente para nuestro curso intensivo en colaboración con la Dra. Maconochie y el Sr. Jerry Wheeler – las discusiones estimulantes que tuvimos en aquellos tiempos y la respuesta recibida de nuestros estudiantes fueron muy útiles para dar forma a estos capítulos.

Isabel dos Santos Silva

Glosario

Algunos conceptos técnicos son difíciles de traducir en términos castellanos adaptables de forma general a las diferentes aceptaciones utilizadas en España y Latino América. Esta situación coincide a menudo con que las expresiones inglesas se han difundido y aceptado ampliamente en los medios epidemiológicos castellano parlantes. El glosario resume los criterios adoptados por el equipo de traducción en estos casos conflictivos y a la vez reúne algunas de las expresiones utilizadas en las diferentes versiones geográficas del castellano.

La primera versión castellana destacada en negrita es la adoptada más frecuentemente en el texto. Por razones de contexto se ha utilizado en ocasiones alguna otra de las aceptaciones listadas. En algunos casos excepcionales se ha mantenido la versión inglesa sin traducir y se ha indicado en caracteres cursivos.

Versión inglesa	Versiones castellanas
Badges	Dosímetros de placa en <i>badges</i>
Bias	Sesgo
Box-and-whisker	Diagrama de cajas, <i>Box-and-whisker</i>
Carry-over effect	Efecto de arrastre
Caveats	Limitaciones
Clinical trial	Ensayo clínico, estudio clínico, ensayo terapéutico
DNA adducts	Aductos de ADN
Effect	Enfermedad, efecto, cáncer
Filter dot hybridization	Hibridación en filtro
Gold Standard	Patrón de oro, estándar de oro
Linkage	Unión, Relación, Confrontación, Encadenamiento, contraste, enlace, cruzamiento
Logrank test	Test de <i>logrank</i>
Matching	Apareamiento
Measurement bias	Sesgo de medida, sesgo de detección, sesgo de medición
Misleading	Engañoso, inducir a error, inducción a error
Modelling	Modelaje, construcción de modelos, modelizar

Odds, Odds Ratio	Odds, Odds Ratio (OR) , Razón de posibilidades, Razón de suertes, Razón de ventajas, Razón de momios, Cociente de productos cruzados, Probabilidades relativas, Cociente entre dos ventajas
Outcome	Resultado , consecuencia, desenlace, enfermedad
Outlier	Valor extremo , valor fuera de rango
Overheads	Gastos institucionales
Overview	Visión general , revisión, meta-análisis
Person-years	Personas-años (p-as)
Proxy	Sustituto , próximo, subrogado
Random	Aleatorio , al azar, aleatorizar
Rate	Tasa
Ratio	Razón
Rate ratio	Razón de tasas (RT)
Record linkage	Unión de registros , relación de registros, contraste de registros
Relative risk	Riesgo relativo (RR)
Screening	Cribaje , tamizaje
Spillover effect	Efecto de arrastre
Survey	Encuesta , estudio
Target population	Población objeto de estudio , población diana
Test	Test , análisis, estudio, prueba
Trial	Ensayo , estudio
Trade off	Balance , equilibrio

Capítulo 1

Introducción a la epidemiología del cáncer

El interés por el cáncer ha aumentado en el último siglo debido al control relativo de las enfermedades infecciosas, a las mejoras sanitarias, a la vacunación y al uso de antibióticos. No obstante, el cáncer en sí no es una enfermedad nueva. En el antiguo Egipto se conocía la existencia del cáncer en humanos. Las autopsias de algunas momias han demostrado la existencia de tumores óseos así como la posibilidad de otros procesos neoplásicos (Brothwell, 1967). En las escrituras médicas chinas y árabes también se describieron síntomas de lo que puede suponerse eran enfermedades malignas. En tiempos de Hipócrates en el siglo cuarto A.C., muchos tipos de tumores estaban reconocidos clínicamente y descritos. Hipócrates introdujo el término carcinoma de *karkinos* (cangrejo): vio el cáncer como un cangrejo en su dispersión por el cuerpo y su persistencia (Long, 1928). Casi 600 años después, Galeno distinguió tres tipos de tumores: «tumores según la naturaleza», que incluían todas las tumefacciones fisiológicas normales como el normal aumento del pecho con la madurez femenina; «tumores que superan la naturaleza», que incluían el proceso productivo tras una lesión, como la proliferación ósea producida durante la recuperación de una fractura; y «tumores contrarios a la naturaleza» que incluían lo que hoy se define como crecimientos neoplásicos así como muchas lesiones inflamatorias (Long, 1928).

Pero no fue hasta el siglo dieciocho que el cáncer empezó a ser estudiado sistemática e intensamente. Bichat (1771-1802) describió la patología de muchas neoplasias en humanos y propuso que el cáncer era una «formación accidental» de tejido formada del mismo modo que cualquier otra porción del organismo. Unas décadas después, Müller (1801-1858) y Virchow (1821-1902) ampliaron sus observaciones utilizando el microscopio para demostrar que el tejido canceroso estaba formado por células (Long, 1928).

Desde entonces, los patólogos y clínicos han considerado los cánceres procedentes de los diferentes órganos del cuerpo, como enfermedades completamente diferentes con morfologías, manifestaciones clínicas y pronósticos diferentes. Pero no es hasta las últimas décadas cuando se pone de manifiesto que sus causas también difieren enormemente. La epidemiología ha sido una de las disciplinas que ha contribuido en mayor parte al estudio de las causas de los diferentes tipos de cáncer y a la evaluación de medidas preventivas.

1.1 ¿Qué es epidemiología del cáncer?

Una definición reciente de epidemiología se indica en el diccionario de epidemiología publicado por Last (1995):

«Epidemiología es el estudio de la distribución y los determinantes de estados o acontecimientos relacionados con la salud, en poblaciones específicas y la aplicación de este estudio en el control de los problemas sanitarios.»

Epidemiología del cáncer es la rama de la epidemiología que se ocupa del estudio de las neoplasias. Esta definición es tan válida para epidemiología del cáncer como lo es para epidemiología en general.

Lo primero a destacar en la definición anterior es que la epidemiología se ocupa de acontecimientos que se producen en poblaciones. Las unidades primarias de interés son grupos de personas, no individuos aislados. Esto es lo que diferencia la epidemiología de la medicina clínica. Los estudios epidemiológicos no sólo se ocupan de los que adquieren la enfermedad, sino también de los que no la adquieren (especialmente, en qué pueden diferir estos dos grupos). Por tanto, a diferencia del médico clínico cuya población de estudio suelen ser pacientes, el epidemiólogo se ocupa de la población de la que proceden los pacientes. Los clínicos dirigen sus preguntas a un paciente en particular: «¿Qué le ocurre a este paciente? ¿Qué puede hacerse para tratarlo?» mientras que el epidemiólogo plantea las mismas preguntas sobre las comunidades enteras: «¿Cuales son las causas principales de muerte o incapacidad en esta población? ¿Qué puede hacerse para reducirlas?» A los clínicos y demás personal sanitario que está acostumbrado a pensar en los problemas de cada paciente en particular, con frecuencia les parece extraño pensar en términos epidemiológicos.

Los estudios de epidemiología del cáncer habitualmente están centrados en la enfermedad «cáncer», aunque no siempre es así. Pueden centrarse en estudios de precursores del cáncer, como por ejemplo la neoplasia intraepitelial cervical (NIC), precursor del cáncer cervical invasivo, o la gastritis atrófica crónica, precursor del cáncer de estómago. Otras veces, el objetivo del estudio puede ser un aspecto relacionado con el cáncer, como el crecimiento o la fertilidad. Por lo tanto, los epidemiólogos no sólo se ocupan de la enfermedad, incapacidad y muerte, sino también de estados de salud y prevención de las enfermedades.

El epidemiólogo se interesa en la distribución de la enfermedad en una población. ¿Qué tipos de personas presentan un mayor riesgo? ¿Cómo cambia la frecuencia de la enfermedad en el tiempo? ¿Cuánto varía de una zona geográfica a otra? Estas preguntas se resumen como *¿quién?*, *¿cuándo?* y *¿dónde?*. La función más básica de la epidemiología del cáncer ha sido simplemente describir la aparición del cáncer humano, observando diferencias, por ejemplo, entre hombres y mujeres, personas de diferente edad, clases socioeconómicas diversas, personas con ocupaciones diferentes, entre diferentes períodos de tiempo, entre diferentes zonas de un país y entre países diferentes. Estas observaciones han sido el punto de inicio en la investigación epidemiológica. Como un detective, el epidemiólogo persigue después las pistas más prometedoras.

El objetivo de la epidemiología no es sólo responder a la pregunta '¿quién? ¿cuándo? y ¿dónde?' sino también a la pregunta '¿por qué?'. ¿Por qué algunas personas o grupos presentan un mayor riesgo que otras? ¿Son las asociaciones entre algunos factores y el riesgo aumentado de la enfermedad causales o *aparentes*? Cuestiones como estas son las preguntas básicas a las que los epidemiólogos intentan responder. Los epidemiólogos consideran su trabajo como la búsqueda de una relación entre dos factores. El primero, denominado 'exposición' (o factor de riesgo), se refiere a cualquier factor que pueda afectar a la salud de la persona, tales como los agentes ambientales (p.e. luz solar, contaminación del aire, agentes ocupacionales), las variables del estilo de vida (p.e. dieta, hábito de fumar, ejercicio físico) o los rasgos constitucionales (p.e. grupo sanguíneo, rasgos genéticos). El segundo es la propia enfermedad (u otro evento sanitario de interés). La función del epidemiólogo es examinar si existe una asociación entre una exposición y una enfermedad y decidir si es probable que la relación observada sea causal: '¿Está el cáncer de pulmón asociado con el hábito de fumar?', '¿Ocasionó la exposición la enfermedad?'

La epidemiología es la única fuente de evidencia científica directa sobre los efectos de la exposición y la posibilidad de prevención de la enfermedad en las poblaciones humanas. Los científicos de laboratorio pueden identificar los compuestos carcinogénicos en el humo del tabaco y pueden producir tumores respiratorios en animales de experimentación forzándolos a inhalar el humo del tabaco (IARC, 1986). No obstante, el argumento de que el fumar causa cáncer de pulmón seguiría no siendo convincente si los epidemiólogos no hubieran demostrado también que el cáncer de pulmón aparece con mayor frecuencia en fumadores que en no fumadores. A diferencia de las ciencias basadas en el laboratorio, la estrategia más frecuente en epidemiología es observar y comparar, más bien que experimentar ya que las consideraciones éticas y prácticas limitan la posibilidad de realizar estudios experimentales en humanos.

Además de establecer si existen asociaciones específicas entre exposición y enfermedad, las investigaciones epidemiológicas se interesan en cuantificar su fuerza. Por lo tanto, una vez encontrada una asociación, la siguiente etapa siempre es cuantificar la magnitud de los posibles efectos. Para hacer esto, los epidemiólogos se plantean preguntas como '¿En cuánto aumenta el consumo de tabaco el riesgo de cáncer de pulmón?' o '¿En cuánto reduce la vacunación contra la hepatitis B el riesgo de cáncer de hígado?'

Uno de los papeles más importantes de la epidemiología es adquirir conocimientos de las causas de la enfermedad o su historia natural, que pueden dar lugar a la introducción de medidas preventivas. La epidemiología puede utilizarse para identificar una causa de enfermedad y un medio de prevención incluso cuando la biología no se conoce completamente. La comprensión de los mecanismos causales no es esencial para aplicar estrategias preventivas eficaces, aunque en el espíritu de la investigación científica aceptado en la cultura occidental, el conocimiento de estos mecanis-

mos sería lo deseable. Podemos reducir considerablemente nuestro riesgo de desarrollar cáncer de pulmón si no fumamos, sin saber cuáles de los componentes de los cigarrillos son específicamente carcinogénicos, ni el mecanismo por el que estas sustancias pueden afectar al control del crecimiento celular.

El epidemiólogo es para la sociedad lo que el médico es para el paciente y la epidemiología es por tanto parte de los fundamentos de la salud pública. La epidemiología describe y determina la aparición de enfermedad en la comunidad. ¿Es la enfermedad (o situación) una prioridad para la intervención sanitaria? ¿Por qué se producen los casos y por qué persiste la enfermedad en la comunidad? La epidemiología también ayuda en la elección de las intervenciones sanitarias con las mejores probabilidades de controlar un problema sanitario particular y ayuda a evaluar su impacto en la comunidad.

1.2 Breve historia de la epidemiología del cáncer

La epidemiología tiene sus orígenes en la idea ya presente en muchas civilizaciones antiguas de que los factores ambientales pueden influir en la aparición de la enfermedad. Hipócrates expresó claramente esta idea aproximadamente hace 2.400 años.

«Quien quiera dedicarse adecuadamente a la ciencia de la medicina debe proceder así. Primero debe considerar qué efectos puede producir cada estación del año; ya que no todas las estaciones son iguales, sino que difieren ampliamente tanto entre ellas como en sus cambios. El siguiente punto son los vientos calientes y el frío, especialmente aquéllos que son universales, pero también aquéllos peculiares para cada región en particular. También debe considerar las propiedades de las aguas (...) y cómo obtienen el agua los nativos, si utilizan aguas pantanosas, blandas o las que son duras procedentes de las alturas rocosas, o saladas y ásperas. El suelo también, si es desnudo y seco o de madera y húmedo, profundo y caliente o alto y frío. También el modo de vida que les gusta a los habitantes, si son muy bebedores, comedores, inactivos o atléticos, laboriosos, si comen mucho y beben poco (...). Utilizando esta evidencia debe examinar los diferentes problemas que aparecen. Ya que si un médico conoce bien estos signos, de preferencia todos, pero en cualquier caso la mayoría, al llegar a una población con la que no está familiarizado, no será ignorante de las enfermedades locales o de la naturaleza de las que aparecen con mayor frecuencia; de modo que no estará perdido en el tratamiento de enfermedades, o equivocarse, como es probable que se produzca si carece de estos conocimientos antes de considerar sus numerosos problemas» (traducido al inglés por Jones (1923)).

A pesar del énfasis en el papel de los factores ambientales en el origen de la enfermedad humana, Hipócrates creía que el cáncer era una enfermedad causada por un exceso de 'bilis negra', que se producía en el bazo y el estó-

magos pero no en el hígado. Sus ideas dominaron la práctica médica durante la Edad Media y no fue hasta la llegada del Renacimiento que una serie de médicos empezaron a discrepar. Ramazzini fue uno de ellos. En el libro titulado *'De Morbis Artificum'* (1713), (traducido al inglés por Wright, 1964), propuso que la elevada aparición de cáncer de mama entre monjas era debido al celibato, una observación que ha superado el paso del tiempo. Rigoni-Stern confirmó posteriormente esta observación en un artículo publicado en 1842 (traducido al inglés por De Stavola (1987)).

En 1775, Percival Pott describió en sus *Observaciones Quirúrgicas* la relación entre el hollín y el cáncer de escroto en deshollinadores. Su trabajo ahora se considera puntal en epidemiología. Pott fue el primero en atribuir una causa ocupacional a esta enfermedad, introduciendo la posibilidad de prevención.

«El destino de esta gente parece singularmente duro; en su infancia con frecuencia son tratados con una gran brutalidad y apenas sobreviven al frío y al hambre; se introducen en lugares estrechos y a veces en chimeneas calientes en las que están enterrados, quemados y prácticamente sofocados; y cuando llegan a la pubertad, son propensos a una desagradable enfermedad dolorosa y fatal» (Pott, 1775).

Henry Butlin, un cirujano del Hospital de San Bartolomeo (Londres) continuó el trabajo de Pott (1845-1912). Waldron (1983) aporta una descripción interesante de sus observaciones:

«Un aspecto de la enfermedad que atrajo a sus estudiosos y que estaba relacionado con su etiología fue que el cáncer del deshollinador parecía ser casi exclusivamente una enfermedad inglesa. Prácticamente no se conocían casos en el Continente, en América o incluso en Escocia. Este problema fue investigado ampliamente por Henry Butlin. Butlin fue al Continente durante el curso de sus investigaciones y, como resultado de reunirse y hablar con deshollinadores, consideró que había establecido las razones de la ausencia de cáncer de escroto entre ellos. En parte era debido al vestuario que llevaban. En 1785, el deshollinador alemán llevaba un traje ajustado completo con la cabeza cubierta [Figura 1.1]. Los trajes no eran anchos y no existían aperturas que permitiesen la entrada del hollín. El deshollinador de Londres en 1851, muestra en cambio un chico con aspecto de granuja, sucio y con blusa ancha y pantalones [Figura 1.2]» (Waldron, 1983).

Butlin se aprovechó de un 'experimento natural' para observar que el vestuario protector (alemán) parece estar asociado a una reducción del riesgo de padecer cáncer de escroto.

El desarrollo y crecimiento del campo de la estadística en el siglo diecinueve permitió estudiar los patrones de la aparición de cáncer en poblaciones específicas. William Farr (1807-1883) en Inglaterra y en colaboración

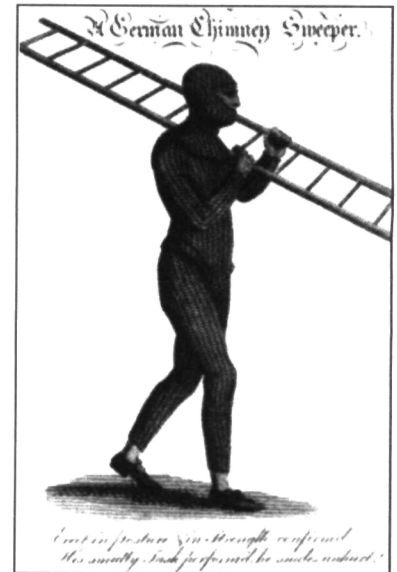


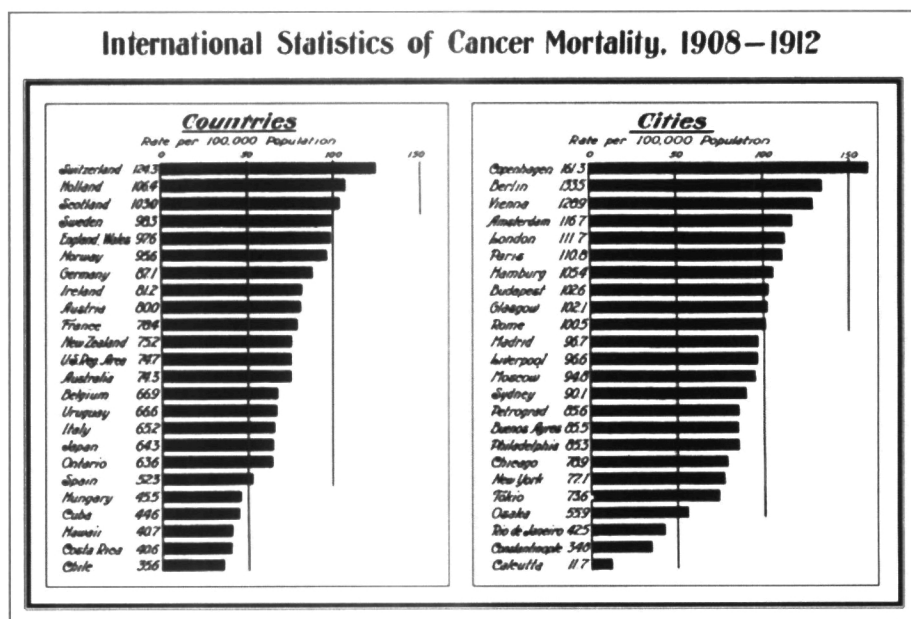
Figura 1.1. Vestuario protector de un deshollinador alemán en 1785 (reproducido con la autorización del BMJ Publishing Group, de Waldron, 1983).



Figura 1.2. Un deshollinador inglés en 1851 (reproducido con la autorización del BMJ Publishing Group, de Waldron, 1983).

con Marc d'Espine en Ginebra, desarrolló un sistema de nomenclatura para la clasificación de enfermedades (Farr, 1975), que formó la base de la *Clasificación Internacional de Enfermedades* (véase Anexo 2.2). La adopción de esta clasificación por muchos países para codificar las causas de muerte registrada en los certificados de defunción mejoró mucho la comparabilidad de las estadísticas internacionales de mortalidad. Uno de los primeros informes sobre las estadísticas de mortalidad por cáncer fue el publicado por Hoffman en 1915 en un libro titulado *La Mortalidad por Cáncer en el Mundo*. Una de las gráficas de este libro se reproduce en la Figura 1.3. Los primeros registros de cáncer basados en la población, que recogen información sobre todos los casos nuevos de cáncer producidos en una población definida, también se establecieron en las primeras décadas de este siglo (Wagner, 1991) (Capítulo 17).

Figura 1.3. Mortalidad por todas las neoplasias malignas para determinados países y ciudades, 1908-1912 (reproducido de Hoffman, 1915).



Aunque antes del siglo veinte se realizaron varios estudios que ahora se considerarían ejemplos de epidemiología del cáncer, el crecimiento de la epidemiología del cáncer (y de la epidemiología en general) no empezó hasta el final de la Segunda Guerra Mundial. En 1954 se publicaron los resultados de dos estudios de cohortes importantes (ahora considerados como *clásicos*). Uno fue el estudio de los médicos británicos (Doll y Hill, 1954), diseñado para investigar la relación del tabaco con el cáncer de pulmón. El otro se diseñó para investigar el riesgo de cáncer de vejiga en la industria química británica (Case *et al.*, 1954; Case y Pearson, 1954).

Estudios de cohortes en poblaciones humanas, especialmente en trabajadores industriales, en pacientes tratados con radioterapia y quimioterapia citotóxica y en víctimas de desastres nucleares, han proporcionado la evidencia más convincente de la relación entre la exposición a agentes específicos y la aparición de cáncer.

En los estudios de cohortes, el epidemiólogo recluta un grupo de personas y reúne información sobre su estado de exposición para identificar a aquellos expuestos a un agente en particular (p.e. fumadores) y a los que no lo están (p.e. no fumadores). A continuación el grupo se sigue en el tiempo y la aparición de la enfermedad (p.e. cáncer de pulmón) en los individuos expuestos se compara con la aparición de la enfermedad en los no expuestos (Capítulo 8). No obstante, el seguimiento y la observación de la aparición de la enfermedad en una población no es una tarea sencilla. Muchas enfermedades (p.e. cáncer) se presentan raramente, de modo que es necesario seguir a grupos poblacionales muy numerosos durante largos períodos de tiempo con objeto de observar casos suficientes con la enfermedad de interés. Esto requiere un gran esfuerzo por parte de los investigadores, cooperación por parte de los sujetos en estudio y sobre todo, un gran presupuesto. Los costes importantes y las dificultades de logística implicadas han favorecido la realización de estudios de cohortes en situaciones en las que la exposición, los registros médicos y las estadísticas vitales se hayan reunido cuidadosamente y estén disponibles para el uso, o donde la sociedad pueda ayudar en los caros esfuerzos requeridos para reunir la información necesaria. A pesar de esto, en países en desarrollo se ha efectuado un número considerable de estudios de cohortes (p.e. Gupta *et al.*, 1984; Geser *et al.*, 1982; Ross *et al.*, 1992). Uno de los más impresionantes fue el seguimiento durante siete años de una cohorte de 42.000 niños de Uganda para investigar el papel etiológico del virus Epstein-Barr en el linfoma de Burkitt (Geser *et al.*, 1982). Este diseño de estudio también se ha utilizado ampliamente en países en desarrollo para examinar los efectos sanitarios (incluyendo el cáncer) de la exposición a un gran número de riesgos ocupacionales (Pearce *et al.*, 1994) y ambientales.

Los problemas de logística y los costes de los estudios de cohortes han propiciado el desarrollo y la popularidad de otro tipo de estudio – el estudio caso-control. En estos estudios, el epidemiólogo compara un grupo de individuos con la enfermedad en estudio ('casos' –p.e. personas con cáncer de pulmón) con un grupo de personas sin esta enfermedad ('controles') para ver si difieren en su historia a las exposiciones (p.e. hábito de fumar) (Capítulo 9). Este tipo de estudio es especialmente adecuado para el estudio del cáncer y otras enfermedades con un período de inducción largo, ya que nos permite retroceder en el tiempo desde los efectos a las causas. Esta secuencia en el tiempo es la inversa de los estudios de cohortes. Esta técnica de investigar de manera retrospectiva es utilizada a diario por los médicos en las historias clínicas. No obstante, los estudios caso-control difieren de las series de casos en cuanto utilizan un grupo control para la comparación con los casos.

Uno de los primeros estudios que utilizaron la estrategia caso-control fue el publicado por Broders (1920) sobre la relación entre el cáncer de células escamosas en el labio y fumar en pipa. Desgraciadamente, no describió el método de selección de los controles. En 1926, Lane-Clayton publicó un estudio caso-control sobre el papel de la historia reproductiva en la etiolo-

gía del cáncer de mama. Este informe es especialmente importante debido a que destacó la necesidad de utilizar un grupo control para la comparación con los casos y comentó los métodos para la selección de los controles hospitalarios en la investigación de hipótesis específicas. Posteriormente, hasta finales de los años 40, no hubo más estudios caso-control de calidad similar al de Lane-Clayton. No obstante, hubo estudios en los que las características de los pacientes con cáncer se compararon con las de un grupo de pacientes sin cáncer; un ejemplo es el estudio de Orr (1933) sobre el mascar betel y los cánceres orales en la India. En 1947, Schrek y Lenowitz publicaron un estudio caso-control que estudia la relación entre el carcinoma de pene y una mala higiene sexual. A principios de los años 50 se publicaron los resultados de cuatro estudios caso-control que relacionaban el hábito de fumar cigarrillos con el cáncer de pulmón (Schrek *et al.*, 1950; Levin *et al.*, 1950; Wynder y Graham, 1950; Doll y Hill, 1950, 1952) y en los años siguientes hubo numerosos trabajos similares sobre muchos cánceres. En los años 50 se desarrollaron numerosos avances metodológicos en relación a los estudios caso-control (p.e. Cornfield, 1951; Mantel y Haenszel, 1959).

El estudio caso-control ha clarificado la comprensión entre asociaciones tales como el primer embarazo en edad tardía y el cáncer de mama (MacMahon *et al.*, 1970); dietilestilbestrol y adenocarcinoma vaginal de células claras en mujeres jóvenes (Herbst *et al.*, 1971); estrógenos exógenos y cáncer de endometrio (Ziel y Finkle, 1975; Smith *et al.*, 1975); consumo de alcohol y tabaco y cáncer de esófago (Tuyns *et al.*, 1977); infección crónica con el virus de la hepatitis B y cáncer de hígado (Prince *et al.*, 1975; Trichopoulos *et al.*, 1978); e infección por el virus del papiloma humano y cáncer de cuello uterino (Muñoz *et al.*, 1992a). El estudio caso-control es especialmente apropiado para la investigación de relaciones causales cuando se dispone de medios relativamente escasos, como en los países en vías de desarrollo.

Un desarrollo más reciente ha sido la aplicación de principios y métodos epidemiológicos al diseño, realización y análisis de ensayos de intervención. En estos estudios los propios investigadores atribuyen a los participantes las exposiciones estudiadas (Capítulo 7). Los experimentos controlados se han convertido en parte habitual de la evaluación de nuevos agentes, métodos preventivos y terapéuticos. Por ejemplo, se han utilizado para evaluar el impacto sanitario del consejo contra el tabaco (Rose y Colwell, 1992); para evaluar el papel de los programas de educación sanitaria en la prevención del cáncer de boca en la India (Gupta *et al.*, 1986); para evaluar los programas de cribaje en el cáncer de mama (p.e. Shapiro *et al.*, 1971); y para evaluar la eficacia de la vacunación contra la hepatitis B en la prevención del cáncer de hígado (p.e. *Gambia Hepatitis Study Group*, 1987).

Es obvio que a partir de la baja supervivencia de muchos cánceres, si queremos tener la situación bajo control, no podemos confiar únicamente en aumentar la disponibilidad de los cuidados médicos. Es vital aumentar la comprensión de los factores genéticos, ambientales y sociales que estimulan la aparición de estas enfermedades, y utilizar este conocimiento para aplicar medidas preventivas eficaces. El objetivo principal de la epide-

miología del cáncer es identificar los factores de riesgo que permitan la introducción prematura de medidas preventivas. No obstante, para conseguir este objetivo, la epidemiología del cáncer requiere una estrategia multidisciplinaria con la colaboración de científicos clínicos, de laboratorio, sociólogos, profesionales de salud pública y demás personal sanitario.

La epidemiología ha contribuido a la prevención de algunos tipos de cáncer bastante antes del conocimiento biológico de la enfermedad. No obstante, los avances recientes en biología están proporcionando nuevas herramientas y orientaciones para las investigaciones epidemiológicas sobre la aparición y distribución del cáncer en las poblaciones. La incorporación de marcadores biológicos en la investigación epidemiológica y el desarrollo de la epidemiología genética son sólo algunos ejemplos de esta estrategia multidisciplinaria.

Lecturas adicionales

* Buck *et al.* (1988) reunieron y comentaron una colección de artículos epidemiológicos clásicos incluyendo muchos de los indicados en el Apartado 1.2.

Cuadro 1.1. Puntos esenciales

- La epidemiología del cáncer se ocupa del estudio de la distribución de la enfermedad 'cáncer' en poblaciones. Su objetivo principal es identificar los factores de riesgo que permitirán la introducción de medidas preventivas.
- La epidemiología del cáncer, y la epidemiología en general, se basa en la comparación de grupos de personas. No obstante, por razones éticas, los métodos epidemiológicos son principalmente observacionales (es decir, no experimentales). El principal reto para los epidemiólogos es identificar y utilizar 'experimentos naturales' que ayudarán a responder la cuestión investigada.
- La epidemiología del cáncer es una ciencia relativamente nueva que ha madurado en la última mitad del siglo veinte. A pesar de su corta edad, ya ha contribuido en gran parte a la comprensión de las causas de los diferentes tipos de cánceres y a la evaluación de medidas preventivas.

Capítulo 2

Medida de exposiciones y resultados

2.1 Introducción

La mayor parte de las investigaciones epidemiológicas implican el estudio de la *relación entre* un tipo de acontecimiento o característica y otro. Considere por ejemplo las siguientes cuestiones:

* *¿Aumenta la ingesta de alcohol el riesgo de cáncer de pulmón?*

Alcohol (exposición) → cáncer de pulmón (resultado)

* *¿Protege la vacunación antihepatitis B contra el cáncer de hígado?*

Vacuna antihepatitis B (exposición) → cáncer de hígado (resultado)

En estas relaciones, se supone que un acontecimiento ('exposición') afecta al otro ('resultado').

La 'exposición' de interés puede estar asociada al aumento o disminución de una enfermedad o a otro evento específico de salud y puede estar relacionada con el ambiente (p.e. contaminación del aire, radón ambiental), estilo de vida (p.e. hábitos de fumar, dieta), o características congénitas o hereditarias (p.e. grupo sanguíneo A, piel clara). El término 'factor de riesgo' con frecuencia se utiliza para describir una variable de exposición.

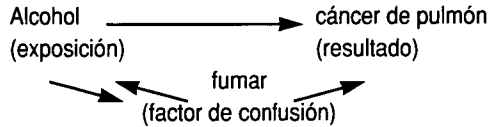
El 'resultado' de un estudio es un término amplio que se refiere a cualquier enfermedad definida, estado de salud, acontecimiento relacionado con la salud, o a la muerte. En algunos estudios puede haber múltiples resultados.

Las exposiciones y los resultados de interés son específicos para el estudio de cada hipótesis y siempre deberían definirse claramente antes de iniciar de un estudio. La exposición de interés en un estudio puede ser el resultado en otro. Por ejemplo, el fumar es claramente la exposición de interés en un estudio que examina si los fumadores tienen más probabilidad de desarrollar cáncer de pulmón que los no fumadores, pero sería el resultado en un estudio que examina la eficacia de un programa de intervención antitabaco en la reducción de la frecuencia de fumar en una población determinada.

En la mayoría de los casos no es suficiente reunir información sólo de la exposición y del resultado de interés. Esto es debido a que su relación puede mezclarse con el efecto de otra exposición con el mismo resultado —estando las dos exposiciones correlacionadas. Esto se conoce como confusión.

Volvamos a considerar la relación entre el consumo de alcohol y el cáncer de pulmón:

* ¿Aumenta la ingesta de alcohol el riesgo de cáncer de pulmón?



Suponga que un investigador observa que el cáncer de pulmón aparece con mayor frecuencia en personas que beben alcohol que en aquellas que no beben. A partir de esta observación no sería posible concluir que la exposición al alcohol aumenta la probabilidad de desarrollar cáncer de pulmón a no ser que el investigador nos pueda convencer de que la relación observada no puede ser debida al hecho de que los bebedores de alcohol fuman más que los no bebedores. En este ejemplo el fumar actúa como un factor de confusión. La confusión puede tratarse en la etapa de diseño y en la etapa de análisis, siempre y cuando se hayan obtenido los datos importantes. Estos aspectos se comentarán con más detalle en los Capítulos 13 y 14.

Por lo tanto, en la mayoría de estudios epidemiológicos es necesario obtener información de tres tipos de variables:

- (1) sobre la(s) exposición(es) primaria(s) de interés,
- (2) sobre otra(s) exposición(es) que pueda(n) influir en el resultado (factores potenciales de confusión) y
- (3) sobre el propio resultado.

Es imposible elegir las mediciones apropiadas para una investigación a no ser que se haya hecho un planteamiento específico y detallado de los objetivos de la investigación. Sin este planteamiento es posible que falte información en variables clave o que se obtengan con poco detalle.

En este capítulo consideraremos diferentes formas de obtener datos sobre las exposiciones y los resultados de interés.

2.2 Tipos de exposición

La gama de exposiciones de interés en epidemiología del cáncer es bastante amplia. Incluye rasgos genéticos (p.e. grupo sanguíneo), variables demográficas (p.e. sexo, edad, raza, nivel socioeconómico), variables reproductoras y relacionadas con la conducta sexual, dieta y constitución corporal, actividad física, hábitos de consumo de tabaco y alcohol, medicaciones anteriores (p.e. uso de anticonceptivos orales), exposiciones medioambientales y ocupacionales, y agentes infecciosos.

Es posible que la característica de interés, la exposición real, no pueda medirse directamente o que sea difícil o imposible de definir. El nivel so-

cioeconómico es un ejemplo de un concepto de este tipo. Los epidemiólogos habitualmente determinan el nivel socioeconómico utilizando variables aproximativas como el tipo de ocupación, el nivel de ingresos, la educación recibida y el lugar de residencia. A su vez, el nivel socioeconómico no es de por sí una causa de enfermedad sino sólo un indicador de nivel o probabilidad de exposición a alguna causa o factor de riesgo que generalmente no se conoce.

2.3 Medida de la exposición

Los datos sobre exposiciones de interés pueden obtenerse mediante entrevistas personales, ya sea cara a cara o por teléfono, cuestionarios auto-administrados, diarios de comportamiento, referencia a registros, análisis biológicos y mediciones en el ambiente. Cuando el sujeto es demasiado joven, está demasiado enfermo o está muerto, también es frecuente obtener datos sobre el mismo a partir de entrevistas a un familiar.

La elección del método de obtención de los datos viene determinada por muchos factores. Los principales: el tipo de estudio a efectuar; el tipo y detalle de los datos requeridos por los objetivos del estudio; la disponibilidad de datos existentes obtenidos para otros fines (p.e. registros); la falta de conocimiento o la mala memoria del sujeto sobre la exposición; la sensibilidad de los sujetos a ser preguntados sobre la exposición; la frecuencia y el nivel de la exposición y su variabilidad en el tiempo; la disponibilidad de métodos físicos o químicos para la determinación de la exposición en el cuerpo humano o en el medio ambiente; y los costes de los diferentes métodos posibles. Los diferentes tipos de datos requieren con frecuencia aproximaciones diferentes para su obtención, pero éstas son también deseables para fines de validación o para reducir errores en la medición (véase Apartado 2.6).

La información obtenida debe incluir detalles sobre la naturaleza exacta de la exposición, su cantidad o dosis y su distribución en el tiempo.

2.3.1 Naturaleza de la exposición

La información obtenida debe ser lo más detallada posible. Por ejemplo, es mejor preguntar sobre las diferentes formas de fumar tabaco por separado (cigarrillos, pipa, puros) que preguntar simplemente sobre 'fumar'. También debe preguntarse sobre el tipo de cigarrillos para obtener información de su contenido de alquitrán. A su vez, debe preguntarse sobre la vía de exposición al agente (p.e. en un estudio de anticonceptivos y cáncer de mama es importante distinguir los anticonceptivos orales de otro tipo de anticonceptivos) y sobre cualquier comportamiento que pueda proteger de la exposición (p.e. en un estudio ocupacional es importante preguntar sobre cualquier comportamiento, como el uso de vestuario protector, que pueda haber protegido a los trabajadores de estar expuestos al factor de riesgo).

2.3.2 Dosis

La exposición pocas veces está simplemente presente o ausente. La mayoría de las exposiciones en las que estamos interesados son variables cuantitativas. Los fumadores pueden clasificarse por el número de cigarrillos consumidos al día, las exposiciones industriales por la magnitud de la exposición (con frecuencia estimadas por la duración del empleo y el tipo de trabajo), las infecciones por la intensidad del proceso infeccioso o por la edad a la exposición, la lactancia por la duración, y las exposiciones psicológicas por alguna escala arbitraria de severidad. Por lo tanto, la simple situación de dos grupos, uno expuesto y uno no expuesto, no es habitual. Las conclusiones de un estudio se refuerzan mucho si existe una 'tendencia' a aumentar la incidencia o la gravedad de la enfermedad al aumentar el nivel de exposición (p.e. una relación entre la exposición y la respuesta).

La dosis puede medirse como la dosis total acumulada (exposición acumulativa), por ejemplo el número total de paquetes de cigarrillos fumados; o como la tasa de exposición, por ejemplo número de cigarrillos fumados al día. Esta última es una dosis por unidad de tiempo.

Es importante considerar que, aunque las determinaciones de la dosis habitualmente se realizan en el ambiente externo del sujeto (p.e. niveles de contaminación ambiental), ésta no es la dosis que importa desde el punto de vista biológico. La dosis biológicamente efectiva es la cantidad de agente externo o su metabolito activo que interacciona con los órganos diana y afecta a sus mecanismos celulares. La mayor parte de las veces esta dosis biológicamente efectiva no puede medirse directamente. No obstante, en algunas circunstancias, puede medirse a través de un estimador – p.e. determinaciones en sujetos humanos de aductos de ADN con nitrosaminas o aflatoxinas. No obstante, estas determinaciones tienen algunos problemas. Por ejemplo, pueden ser buenos marcadores de exposiciones actuales o recientes pero no de exposiciones en el pasado (Apartado 2.4.4).

2.3.3 Tiempo

En la medida que sea posible, cada exposición debe caracterizarse en relación a cuando apareció por primera vez, cuando finalmente terminó (si procede) y cómo se distribuyó durante el periodo en que actuó (¿fue periódica o continua? ¿varió la dosis con el tiempo?). Deberían obtenerse detalles similares para cualquier comportamiento que pueda proteger frente a la exposición.

Existe un periodo de tiempo limitado (el *intervalo de tiempo crítico*) en el que el agente podría haber causado el cáncer. Lamentablemente, el inicio y el final de este intervalo de tiempo crítico no se conocen y es probable que su longitud varíe de un individuo a otro. La obtención de datos relacionados con el tiempo de la exposición permitirá, como mínimo, realizar análisis para examinar su posible magnitud. Estos análisis pueden incluir el estudio de los efectos del tiempo desde la primera exposición y del tiempo desde la última exposición.

El patrón de exposición puede ser también importante. Frente a una ex-

posición total similar, los efectos pueden ser diferentes si se acumulan en un corto espacio de tiempo o se reparten a lo largo de un periodo prolongado con intensidades reducidas (p.e. una exposición al sol constante frente a una intermitente; la exposición crónica a niveles bajos de radiación ionizante frente a la exposición aguda a niveles altos).

2.4 Fuentes de datos de exposición

2.4.1 Cuestionarios

Los cuestionarios se utilizan para obtener datos de exposición en estudios epidemiológicos planteando la misma serie de preguntas a cada participante del estudio de manera estandarizada. Los cuestionarios pueden ser auto-administrados o administrados por un entrevistador.

Los objetivos de un cuestionario de investigación son obtener, con un error mínimo, determinaciones de las variables de exposición de interés para el estudio. Por lo tanto, las preguntas a incluir en un cuestionario deben referirse directamente a los objetivos del estudio. Algunos de los principios básicos que pueden tenerse en cuenta en el diseño de un cuestionario se comentan en el Anexo 2.1. Todo cuestionario debe probarse previamente en una muestra de sujetos de la población a estudiar para asegurar que las preguntas se entienden y proporcionan respuestas apropiadas.

Cuestionarios auto-administrados

Los cuestionarios auto-administrados se utilizan para estudiar sujetos a los que se les pide que los completen ellos mismos. Pueden distribuirse y devolverse por correo si es factible y más conveniente. Son especialmente apropiados cuando se requieren pequeñas cantidades de datos razonablemente simples o para documentar un comportamiento sensible o socialmente indeseable. Estos cuestionarios son una de las maneras más baratas de obtener información. Una de las principales limitaciones es que sólo pueden utilizarse en poblaciones con un cierto nivel cultural. También existe un control relativamente bajo por parte del investigador sobre la calidad de los datos obtenidos.

Entrevistas personales (cuestionarios administrados por un entrevistador)

El uso de entrevistadores puede mejorar la calidad del estudio aumentando la participación, motivando al sujeto a responder adecuadamente e intentando obtener datos más completos. No obstante, los entrevistadores también pueden comprometer los resultados si influyen directa o indirectamente las respuestas del sujeto.

Una entrevista es una conversación entre el entrevistador y el entrevistado y por tanto es esencial que los entrevistadores establezcan una buena relación con los entrevistados desde el inicio de la entrevista. Los sujetos del estudio deben confiar en los entrevistadores y por tanto éstos deben elegirse considerando las normas culturales de la población donde se va a reali-

zar el estudio. Por ejemplo, en algunas sociedades no se permitiría a los entrevistadores hombres que entrevistaran a mujeres. El nivel cultural de los entrevistadores puede también influir en el grado de participación y/o en la exactitud de la información proporcionada por los entrevistados. Es esencial que el entrevistado sienta que el entrevistador le entiende y que no existen 'barreras de comunicación'.

Las entrevistas cara a cara son claramente mejores para la obtención de grandes cantidades de datos complejos. No obstante, cuando los sujetos están muy dispersados geográficamente y el cuestionario es breve, las entrevistas telefónicas pueden ser una mejor alternativa. Esta estrategia sólo es factible en poblaciones donde exista el uso difundido de teléfonos, lo que no es el caso de muchas zonas del mundo. Incluso en sociedades donde existe el uso difundido de teléfonos, debe considerarse que ciertos grupos de personas quedarán sistemáticamente excluidos del estudio porque no tienen teléfono (p.e. personas sin hogar, poblaciones nómadas) o porque no quieren proporcionar información personal de este modo.

Los próximos o sustitutos que responden son personas que proporcionan información sobre la exposición en lugar de los propios sujetos incluidos en el estudio (sujetos índice). Se utilizan en epidemiología cuando los sujetos índice por alguna razón (p.e. muerte, demencia) no son capaces de proporcionar los datos requeridos. Los estudios que implican a niños también tienen que confiar en los próximos. Los próximos o sustitutos normalmente proporcionan información menos válida que los sujetos índice. Por ejemplo, con frecuencia tienden a infraestimar las exposiciones ocupacionales y a decir que el sujeto índice tiene un trabajo mejor del que tiene realmente. La familiaridad del entrevistado con el sujeto del estudio es un determinante importante de la calidad de información obtenida por este método; en general, las esposas y madres en el caso de niños, tienden a proporcionar una información más exacta que otros miembros de la familia.

2.4.2 Diarios

Los diarios son registros detallados de la exposición, actualizados por el mismo sujeto. Generalmente son de respuesta abierta y tienen forma de cuadernillo en el que el sujeto anota cada ocurrencia de un comportamiento especial como ejercicio físico, consumo de alcohol, ingesta dietética, actividad sexual, medicación, etc. Se supone que los diarios son muy exactos en la determinación del comportamiento actual debido a que no dependen de la memoria. También permiten la obtención de más detalles sobre la exposición que mediante el cuestionario. Por ejemplo, el sujeto puede pesar los alimentos antes de consumirlos.

La principal limitación del uso de diarios es que sólo pueden medirse exposiciones actuales. Además, los diarios generalmente requieren más tiempo y habilidad por parte de los sujetos que los otros métodos, de modo que el cumplimiento puede ser un problema. La formación de los sujetos en las habilidades necesarias para llevar un diario exacto puede requerir mucho

tiempo para los sujetos y el personal del estudio. Por esta razón se utilizan raramente en países donde el analfabetismo es un problema.

2.4.3 Registros

Los datos sobre la exposición de interés pueden obtenerse a partir del censo, registros de empleo, médicos (paciente hospitalizado o ambulatorio), registro de cáncer, certificado de nacimiento y certificado de defunción. Normalmente, como la obtención de los datos ya se ha efectuado con otros fines diferentes a la investigación epidemiológica, el investigador no tiene control sobre las variables que se registraron, cómo se plantearon las preguntas. Los datos de los registros están generados con frecuencia por muchas personas con una formación poco uniforme. La disponibilidad y calidad de los registros en muchos países tiende a ser baja.

A pesar de estas limitaciones, existen varias ventajas para el uso de registros sobre los otros métodos de obtención de datos. Los costes del estudio suelen ser bajos y el tiempo del estudio se reduce porque parte o toda la obtención de datos ha sido realizada por otros. También ofrecen un potencial de datos casi completos en una población bien definida y puede obtenerse información sin la necesidad de contactar con los sujetos o sus parientes. La exactitud de algunos tipos de datos (p.e. toma de medicación, exposición ocupacional) puede ser mejor que la obtenida a través de una entrevista personal. Pueden, por ejemplo, evitar errores debidos a falta de memoria o a un desconocimiento completo de la exposición de interés por parte del sujeto.

Las características y limitaciones de algunos sistemas de recogida de datos rutinarios se comentarán con detalle más adelante en el Apartado 2.9 y en el Capítulo 11.

2.4.4 Determinaciones biológicas

Las determinaciones efectuadas directamente en el cuerpo humano o en sus productos representan, en principio, el método ideal para medir la exposición. Las determinaciones biológicas tienden a ser más objetivas ya que son independientes de las percepciones del sujeto e independientes del investigador si se utilizan métodos instrumentales o de laboratorio. También pueden estar más próximas a la dosis biológicamente efectiva, es decir, el nivel que interfiere realmente con las células en el órgano(s) diana.

La aplicación epidemiológica de las medidas de la exposición en el cuerpo humano ha recibido recientemente una atención especial con el desarrollo de técnicas de laboratorio cada vez más refinadas destinadas a determinar los metabolitos activos de carcinógenos y sus productos de interacción con el ADN o proteínas (aductos), de tal manera que se ha popularizado un nuevo término «epidemiología molecular» para describir estudios epidemiológicos que incorporan un componente importante de resultados basados en la tecnología biológica moderna.

Un ejemplo interesante de la aplicación con éxito de la epidemiología molecular es la determinación de las aflatoxinas en la dieta. Esta toxina se

produce por el hongo *Aspergillus flavus* que crece en alimentos almacenados como los cacahuets en climas tropicales, especialmente en el Este Asiático y en el África subsahariana. Es una potente causa experimental de cáncer de hígado en animales de laboratorio. No obstante, la mayor parte de la investigación epidemiológica realizada en el pasado estaba obstaculizada por la dificultad de determinar la cantidad de aflatoxinas consumida habitualmente por un individuo. Recientemente se han descubierto marcadores biológicos para la estimación del consumo individual actual o reciente de aflatoxinas, los cuales miden metabolitos de aflatoxinas y aductos de ADN en orina. Estos marcadores se utilizaron en un estudio realizado en Shanghai, China (Qian *et al.*, 1994), en el que se relacionó la incidencia de cáncer de hígado en aproximadamente 18.000 hombres chinos con las determinaciones urinarias de su exposición a aflatoxinas. Los resultados de este estudio han proporcionado la evidencia más directa hasta ahora de un papel etiológico de las aflatoxinas en la carcinogénesis hepatocelular humana. No obstante, este marcador biológico todavía está lejos de ser ideal, ya que no mide la exposición en el pasado, la cual puede ser muy importante en el estudio de la influencia de esta sustancia química en la patogénesis del cáncer de hígado.

También se han desarrollado nuevos ensayos de laboratorio para ayudar a la evaluación de la exposición a agentes infecciosos tales como el Virus del Papiloma Humano (VPH) (Muñoz *et al.*, 1992b) y *Helicobacter pylori* (IARC, 1994a), que han ayudado a clarificar su papel en la etiología de los cánceres de cuello uterino y de estómago, respectivamente.

El uso de análisis biológicos en estudios epidemiológicos depende principalmente de la disponibilidad de un ensayo adecuado, de la factibilidad (p.e. disponibilidad del equipo de laboratorio) y de los costes. Además, la mayoría de estas determinaciones están limitadas por el hecho de que sólo pueden evaluar exposiciones actuales, que generalmente no son las más importantes en epidemiología del cáncer. Son especialmente útiles para atributos que permanecen estables (p.e. rasgos genéticos). Los bancos de muestras biológicas son una excepción. Las muestras biológicas tomadas en el pasado (mucho antes de que los sujetos desarrollen el resultado de interés) pueden ser recuperadas y analizadas. Por ejemplo, pueden tomarse las muestras de sangre y de orina de todos los miembros de una cohorte determinada, en el momento de su inclusión en el estudio y conservarlas congeladas. Estas muestras pueden volverse a analizar más adelante si se dispone de técnicas más actualizadas.

2.4.5 Determinaciones medioambientales

Las determinaciones medioambientales incluyen determinaciones de agentes existentes en el aire (p.e. contaminantes del aire, polvo), agua (p.e. flúor), suelo (p.e. elementos químicos), alimentos (p.e. composición de nutrientes), etc. Estas determinaciones pueden realizarse en hogares, lugares de trabajo, lugares de recreo o en el medio ambiente en general. Son especialmente útiles cuando los sujetos no son conscientes de la exposición (p.e. niveles internos de radiación) o no la pueden recordar bien.

El valor de estas medidas depende del método de determinación utilizado y de la toma de muestras del ambiente. Lo ideal es que la determinación de los agentes medioambientales se realice para cada sujeto en estudio a lo largo del periodo etiológicamente importante. De este modo reflejarán lo mejor posible un atributo personal. Por ejemplo, la medida individual de la exposición a una radiación ionizante puede realizarse solicitando a cada sujeto en estudio que lleve una placa grabadora durante el periodo de estudio; la ingesta individual de nutrientes puede medirse analizando químicamente porciones idénticas de todos los alimentos y bebidas consumidas por el sujeto durante el periodo de estudio. No obstante, en la mayor parte de los casos, esta estrategia no es posible debido a limitaciones de tiempo y coste, a aspectos técnicos y a la falta de cumplimiento por parte del sujeto. Frecuentemente debemos confiar en las determinaciones efectuadas a una muestra de sujetos en estudio sólo en ciertos puntos en el tiempo. La elección de la muestra en la cual se realizarán las determinaciones y el tiempo en las que se llevarán a cabo es obviamente crítica para la validez de las determinaciones.

Una de las limitaciones de las determinaciones ambientales es que normalmente reflejan niveles actuales de exposición. En determinadas situaciones, puede ser razonable suponer que las concentraciones actuales de agentes del medio ambiente están altamente correlacionadas con los niveles de exposición en periodos etiológicamente importantes del pasado. Puede disponerse de registros de las determinaciones efectuadas en el pasado, pero deben utilizarse con precaución ya que se realizaron con otros fines diferentes al estudio epidemiológico utilizando métodos que ahora pueden considerarse inadecuados. Cuando no se dispone de determinaciones ambientales anteriores, pueden utilizarse medidas aproximadas de las exposiciones en el pasado. Por ejemplo en un estudio ocupacional, la información sobre el 'tipo de trabajo', 'año de empleo' y 'duración del empleo' puede utilizarse para clasificar a los trabajadores en función de su estado de exposición. Esta información puede extraerse de los registros de empleo o mediante cuestionarios.

2.5 Medida del resultado

Al igual que con la medida de la exposición, los datos sobre el evento de interés pueden obtenerse a partir de diferentes fuentes. Pueden enviarse cuestionarios (o realizar llamadas de teléfono) regularmente a los sujetos en estudio para evaluar su estado de salud. Pueden realizarse entrevistas personales periódicas con controles clínicos que pueden incluir medidas biológicas y otros métodos de diagnóstico apropiados (p.e. radiografía, endoscopia, ultrasonidos, etc.). De manera alternativa, la información sobre los resultados de interés, y en particular sobre la aparición de cáncer, puede obtenerse a partir de registros tales como los registros hospitalarios, registros de cáncer, certificados de defunción o algún otro método especializado de vigilancia (véase Apartado 2.9). En este caso, la identificación de los casos está limitada a los resultados que se registran rutinariamente, al grado de su cumplimiento y a la manera como están codificados.

Debido a que las enfermedades neoplásicas se desarrollan tras un periodo de latencia generalmente prolongado y son relativamente raras, los estudios sobre la relación entre exposiciones carcinogénicas sospechadas y cáncer pueden requerir observar a los participantes durante un largo periodo de tiempo. Una manera de reducir este periodo es utilizar variables intermedias substitutivas de cáncer, es decir, utilizar como resultado un acontecimiento biológico que se cree se encuentra en la vía causal entre la exposición y el cáncer. Estos estudios en principio son más rápidos, más pequeños y menos caros que los estudios que utilizan la enfermedad como resultado final. Por ejemplo, un estudio de la dieta en relación con el metabolismo de estrógenos podría efectuarse en varias docenas de pacientes mientras que un estudio de intervención de dieta y cáncer de mama como variable final requeriría decenas de miles de mujeres con muchos años de seguimiento (Schatzkin *et al.*, 1990). La suposición subyacente en estos estudios es que la relación observada entre exposición (p.e. dieta) y la variable intermedia (p.e. metabolismo de estrógenos) refleja una relación similar entre la exposición y el cáncer de interés. Por supuesto, esta relación tiene que validarse antes de que cualquier variable intermedia pueda utilizarse como variable substitutiva de cáncer (Toniolo *et al.*, 1997).

2.6 Validez y fiabilidad de las medidas de exposición y resultado

2.6.1 Validez

La validez se define como la capacidad de un instrumento (por ejemplo, un cuestionario o una prueba de laboratorio) para medir lo que intenta medir. La validez sólo puede determinarse si existe un procedimiento de referencia o 'estándar de oro': un procedimiento considerado definitivo para establecer si alguien tiene la característica de interés. Por ejemplo, la información sobre el peso al nacer obtenida en una entrevista puede validarse frente a anotaciones obstétricas del hospital, los cuestionarios de frecuencia de consumo de alimentos pueden confrontarse a diarios de alimentación o a determinaciones biológicas. Lamentablemente, en algunas circunstancias no existe un procedimiento de referencia obvio y el mejor método disponible debe aceptarse como el estándar.

Consideremos el ejemplo simple de una prueba que sólo pueda dar un resultado positivo o negativo (p.e. binario). Cuando los sujetos de la muestra se hayan examinado mediante la prueba en estudio y el estándar de oro, los resultados pueden expresarse en una tabla de 2 x 2 como se muestra en la Tabla 2.1.

La sensibilidad de la prueba en estudio es la proporción de los individuos clasificados como positivos por el estándar de oro que se identifican correctamente por la prueba en estudio:

$$\text{Sensibilidad} = a/(a+c)$$

		'Estándar de oro'	
		Positivo	Negativo
Prueba en estudio	Positivo	a	b
	Negativo	c	d

a, verdaderos positivos; b, falsos positivos; c, falsos negativos; d, verdaderos negativos.

Tabla 2.1. Representación general de una tabla de 2 x 2 para evaluar la validez de una prueba en la que sólo cabe un resultado binario.

La especificidad de la prueba en estudio es la proporción de los individuos clasificados como negativos por el estándar de oro que se identifican correctamente por la prueba en estudio:

$$\text{Especificidad} = d/(b+d)$$

El valor predictivo de un resultado positivo representa la probabilidad de que alguien con un resultado positivo en la prueba en estudio tenga realmente la característica de interés, determinada por el patrón de oro:

$$\text{Valor predictivo de un resultado positivo} = a/(a+b)$$

El valor predictivo de un resultado negativo representa la probabilidad de que alguien con un resultado negativo en la prueba en estudio no tenga la característica de interés, determinada por el patrón de oro:

$$\text{Valor predictivo de un resultado negativo} = d/(c+d)$$

Una prueba ideal tiene una alta sensibilidad (identifica correctamente a una alta proporción de individuos realmente expuestos o enfermos) y una alta especificidad (da pocos resultados positivos en individuos no expuestos o no enfermos). En el Ejemplo 2.1, la prueba ViraPap® tenía sensibilidad y especificidad altas, lo que indicaba que la prueba era muy válida en la detección de la infección del cuello uterino por VPH (en comparación con la prueba de hibridación *Southern*) y por lo tanto sus resultados se verían poco afectados por el error de medición.

El valor predictivo de un resultado depende en gran parte de lo común que sea la enfermedad (u otra característica de interés), mientras que la sensibilidad y la especificidad no se afectarán considerablemente. Cuando cambia la frecuencia de la enfermedad, el número de personas enfermas determinado por el 'estándar de oro' (columna de la izquierda) cambia en proporción al número de personas no enfermas (columna de la derecha). A diferencia de la sensibilidad y la especificidad, el valor predictivo de un resultado positivo de la prueba ViraPap® depende de los números en las dos columnas y cambiará si cambia la frecuencia de la enfermedad.

En el Ejemplo 2.2, el valor predictivo de una prueba ViraPap® positiva disminuyó considerablemente (del 74% al 45%). Esto es debido a que la proporción de mujeres infectadas por VPH (determinada por el estándar de oro) fue mucho mayor ($69/450 = 15\%$) en la muestra de mujeres que asistieron a la clínica de enfermedades de transmisión sexual (Tabla 2.2) que entre el grupo de mujeres aparentemente sanas ($23/450 = 5\%$). Así pues, las pruebas de diagnóstico útiles en medicina clínica pueden ser poco eficaces en estudios epidemiológicos o en programas de estudio poblacionales. En medicina clínica, las pruebas de diagnóstico se aplican a una selección de pacientes con una aparición alta de la condición de interés. En esta situación, la prueba puede tener un valor predictivo alto. En un estudio epide-

Ejemplo 2.1. Se ha desarrollado una serie de métodos de laboratorio para la detección de la infección por virus del papiloma humano (VPH) en el cuello uterino. En un estudio realizado hace unos años se evaluó la eficacia de una prueba nueva de hibridación filter dot (ViraPap®) disponible en el mercado, comparando sus resultados con los obtenidos utilizando un 'estándar de oro' en una muestra de 450 mujeres consecutivas que asistieron a una clínica de enfermedades de transmisión sexual en el estado de Washington (EE.UU.) durante los años 1987-88 (Kiviat et al., 1990). La prueba de hibridación por transferencia Southern, que es muy cara y requiere mucho tiempo, se tomó como el 'estándar de oro' en este estudio. Los resultados se muestran en la Tabla 2.2.

Tabla 2.2.

Comparación entre los métodos Vira-Pap® y de hibridación por transferencia Southern en el diagnóstico de la infección del cuello uterino por VPH en una muestra de mujeres que asistieron a una clínica de enfermedades de transmisión sexual^a.

		Hibridación por transferencia Southern (prueba de 'estándar de oro')		
		Positivo	Negativo	Total
ViraPap® (prueba nueva)	Positivo	62	22	84
	Negativo	7	359	366
	Total	69	381	450

^a Modificado de Kiviat et al., 1990

Estos datos proporcionan lo siguiente para la prueba ViraPap®:

Sensibilidad = $62/69 = 90\%$

Especificidad = $359/381 = 94\%$

Valor predictivo de una prueba ViraPap® positiva = $62/84 = 74\%$

Valor predictivo de una prueba ViraPap® negativa = $359/366 = 98\%$

miológico de una población no seleccionada, la misma prueba puede tener un valor predictivo bajo porque la frecuencia de la condición es mucho menor. Por ejemplo, la mamografía tiene un valor predictivo alto como prueba diagnóstica para el cáncer de mama en mujeres que consultan a los médicos por un bulto en el pecho, pero un valor predictivo bajo cuando se utiliza para estudiar mujeres aparentemente sanas de la población. Estos aspectos se comentarán más adelante en el Capítulo 16.

La selección de un estándar de oro es un aspecto crucial en la evaluación de la validez de una determinación. Lamentablemente, en muchas circunstancias no existe un estándar de oro apropiado y se está forzado a confiar en el mejor método disponible en aquel momento. Por ejemplo, el método de hibridación por transferencia Southern se consideró como el estándar de oro en la detección de la infección del cuello uterino por VPH durante muchos años. El desarrollo en los últimos años de los métodos de reacción en cadena de la polimerasa (PCR) para amplificar secuencias de ADN específicas de VPH, ha convertido a este nuevo método en el nuevo 'estándar de oro'.

Ejemplo 2.2. Suponga que la prueba ViraPap® considerada en el ejemplo anterior se utilizó después en una muestra de 450 mujeres aparentemente sanas que visitaron a sus médicos de cabecera para una revisión médica. Los resultados se presentan en la Tabla 2.3.

		Hibridación por transferencia <i>Southern</i> (prueba de 'estándar de oro')		
		Positivo	Negativo	Total
ViraPap® (prueba nueva)	Positivo	21	26	47
	Negativo	2	401	403
	Total	23	427	450

Para la prueba ViraPap® se puede calcular lo siguiente:

Sensibilidad = $21/23 = 91\%$

Especificidad = $401/427 = 94\%$

Valor predictivo de una prueba ViraPap® positiva = $21/47 = 45\%$

Valor predictivo de una prueba ViraPap® negativa = $401/403 = 100\%$

Tabla 2.3.

Comparación de los métodos ViraPap® y de hibridación por transferencia *Southern* en la detección de la infección del cuello uterino por VPH en mujeres aparentemente sanas: datos hipotéticos.

En el Ejemplo 2.3 la validez de la prueba ViraPap® (medida por su sensibilidad y especificidad) es ahora mucho menor que cuando se utilizó la técnica de hibridación por transferencia *Southern* como estándar de oro (Ejemplo 2.1). Esto es debido a que el método PCR es más sensible y más específico que la técnica de hibridación por transferencia *Southern*.

Algunas pruebas proporcionan resultados que no son una simple respuesta si/no. En su lugar, proporcionan valores numéricos en una escala continua de medición. En estas situaciones, se obtiene una alta sensibilidad al precio de una baja especificidad y viceversa. Por ejemplo, cuanto más alta es la presión sanguínea, más probable es la existencia de una enfermedad hipertensiva. Si el nivel diagnóstico o de cribaje para la hipertensión se fija en, por ejemplo, una presión sanguínea diastólica de 90 mmHg, se detectaría la mayoría de los pacientes hipertensos (alta sensibilidad), pero muchos pacientes no enfermos (con una presión sanguínea diastólica superior a 90 mmHg) se clasificarán erróneamente como hipertensos (baja especificidad). Si el nivel de cribaje para la enfermedad hipertensiva se fija a 110 mmHg para la presión sanguínea diastólica, la mayoría de los individuos no hipertensos se excluirían de la atención posterior (alta especificidad), pero se dejarían de identificar muchos pacientes hipertensos (con presiones sanguíneas diastólicas menores de 110 mmHg) (baja sensibilidad).

En el Ejemplo 2.4 podían haberse tomado otros valores de concentración sanguínea como valores límite para definir los resultados del ensayo como 'positivos' o 'negativos'. En la Tabla 2.6 se presentan los valores de sensibilidad y especificidad del análisis de sangre para diferentes valores límite.

Ejemplo 2.3. La utilidad de ViraPap® se comparó con la de la reacción en cadena de la polimerasa (PCR) en 373 mujeres con un diagnóstico reciente de cáncer de cuello uterino. Los resultados se muestran en la Tabla 2.4.

Tabla 2.4. Comparación de los métodos ViraPap® y de la reacción en cadena de la polimerasa (PCR) en la detección de infección del cuello uterino por VPH^a.

		PCR (prueba de 'estándar de oro')		
		Positivo	Negativo	Total
Prueba ViraPap®	Positivo	163	11	174
	Negativo	120	79	199
	Total	283	90	373

^a Muñoz *et al.* (no publicado)

Para la prueba ViraPap® se puede calcular lo siguiente:

Sensibilidad = $163/283 = 58\%$

Especificidad = $79/90 = 88\%$

En la Tabla 2.6, la sensibilidad de la prueba de laboratorio disminuyó al aumentar el valor límite mientras que se produjo lo contrario con la especificidad. Esto se ilustra claramente en la Figura 2.1.

Una manera de resumir la validez de una determinación continua es representar gráficamente la sensibilidad frente a (1 - especificidad) para dife-

Ejemplo 2.4. Se ha desarrollado una prueba de laboratorio nueva que mide la concentración de una enzima particular en la sangre y queremos evaluar su valor en el diagnóstico de un cáncer en particular. Esta prueba nueva se aplicó a 360 pacientes hospitalizados y los resultados se compararon con los de un examen anatómo-patológico. Las concentraciones sanguíneas ≥ 40 UI se consideraron como valores 'positivos'. Los resultados se muestran en la Tabla 2.5.

Tabla 2.5. Comparación de una prueba de laboratorio nueva con los exámenes anatómo-patológicos en el diagnóstico de un cáncer determinado: datos hipotéticos.

		Examen anatómo-patológico (prueba de 'estándar de oro')		
		Positivo	Negativo	Total
Análisis de sangre	Positivo (≥ 40 IU)	190	80	270
	Negativo (< 40 IU)	0	90	90
	Total	190	170	360

Para la prueba de laboratorio nueva se puede calcular lo siguiente:

Sensibilidad = $190/190 = 100\%$

Especificidad = $90/170 = 53\%$

Valor límite (IU)	Resultado del análisis de sangre	Resultado del examen anatómo-patológico	Número de pacientes
40 ^a	+	+	190
	+	-	80
	-	+	0
	-	-	90
80 ^a	+	+	188
	+	-	42
	-	+	2
	-	-	128
120 ^a	+	+	173
	+	-	25
	-	+	17
	-	-	145
280 ^a	+	+	95
	+	-	0
	-	+	95
	-	-	170

^a Los resultados del análisis de sangre iguales o superiores al valor límite se consideraron positivos: +, resultado positivo; -, resultado negativo.

40 IU:	sensibilidad = 190 / 190 = 100%	especificidad = 90 / 170 = 53%
80 IU:	sensibilidad = 188 / 190 = 99%	especificidad = 128 / 170 = 75%
120 IU:	sensibilidad = 173 / 190 = 91%	especificidad = 145 / 170 = 85%
280 IU:	sensibilidad = 95 / 190 = 50%	especificidad = 170 / 170 = 100%

Tabla 2.6.
Sensibilidad y especificidad del análisis de sangre para diferentes valores límite: datos hipotéticos.

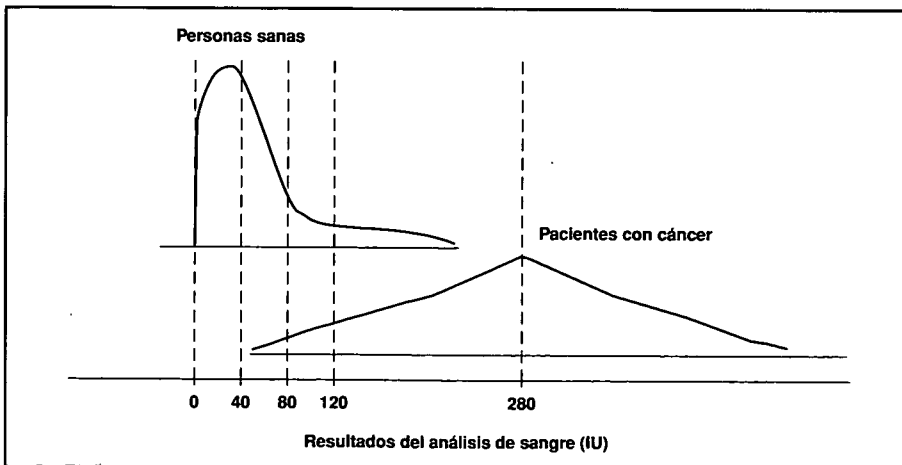


Figura 2.1.
La curva superior describe la distribución de los resultados del análisis de sangre entre individuos sanos y la curva inferior la distribución entre pacientes con cáncer (definida por el examen anatómo-patológico). Se utilizaron diferentes valores límite para clasificar los resultados del análisis de sangre como 'positivos' o 'negativos'.

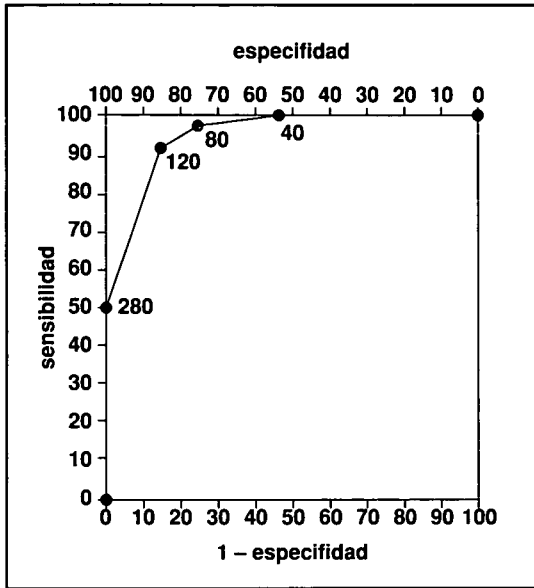


Figura 2.2.
Curva característica operativa del receptor (COR) construida con los datos de la Tabla 2.6 y la Figura 2.1.

rentes valores límite. Esta curva se denomina *curva característica operativa del receptor* (COR) [Receiver Operating Characteristic (ROC)]. La curva COR correspondiente al ejemplo anterior se muestra en la Figura 2.2.

Cuanto más próxima esté la curva COR de una determinada prueba del extremo superior izquierdo de la Figura 2.2, donde se aumentan la sensibilidad y la especificidad, más válida será la prueba. Una prueba con una curva próxima a la diagonal será inútil y no mejor como técnica diagnóstica que la mera suposición.

2.6.2 Fiabilidad

La fiabilidad, a veces denominada también repetibilidad o reproducibilidad, es una medida de la consistencia de la eficacia de un instrumento cuando se utiliza repetidamente en condiciones similares. La fiabilidad de una medición es una condición necesaria pero no suficiente para la validez. En otras palabras, una prueba puede proporcionar el mismo resultado de manera consistente, pero el resultado puede no ser el real (válido). La baja fiabilidad de una medición puede ser debida a la variación entre sujetos en diferentes ocasiones (variación biológica) o a errores en la técnica de medición (variación del observador y del instrumento). Normalmente es acertado incluir algunos controles de repetibilidad de las determinaciones de las principales exposiciones y eventos finales en un estudio epidemiológico. Estos pueden ser medidas de:

(1) Fiabilidad intra-observador o intra-determinación

Se puede determinar con el mismo observador efectuando las mismas mediciones a los mismos sujetos en dos o más ocasiones diferentes. Por ejemplo, los datos de los registros médicos pueden extraerse dos veces por la misma persona en dos ocasiones diferentes; el mismo entrevistador puede volver a entrevistar a los sujetos después de un determinado intervalo de tiempo; muestras biológicas en duplicado pueden volver a procesarse por el mismo técnico de laboratorio en tiempos diferentes. Las dos (o más) determinaciones separadas se comparan. La especificación del intervalo de tiempo adecuado entre medidas varía según el tipo de determinación del evento final o de la exposición. Si es demasiado corto, los sujetos y/o observadores pueden recordar el resultado anterior; si el intervalo de tiempo es demasiado largo, puede haber cambiado el estado de exposición real o evento final del paciente (no es un problema cuando los datos se extraen de registros médicos).

(2) Fiabilidad inter-observador

La fiabilidad inter-observador se puede evaluar en los mismos sujetos determinados por dos o más observadores independientes. Por ejemplo, la fiabilidad de dos o más analistas puede compararse solicitándoles por separa-

do que extraigan información de registros médicos de los mismos sujetos; o la fiabilidad de dos o más entrevistadores puede compararse solicitándoles que entrevisten a los mismos sujetos en dos ocasiones diferentes. Al igual que antes, el intervalo entre medidas requiere una cuidadosa consideración.

Consideremos el ejemplo simple de una prueba que sólo puede dar un resultado positivo o negativo (es decir, binario). La concordancia entre pares de determinaciones efectuadas por dos observadores independientes en los mismos sujetos puede establecerse en una tabla de 2×2 (Tabla 2.7).

		Observador B		Total fila
		Positivo	Negativo	
Observador A	Positivo	a	b	a+b
	Negativo	c	d	c+d
Total columna		a+c	b+d	
		Total final		N

Tabla 2.7.

Representación general de una tabla para evaluar la fiabilidad entre dos observadores para una prueba binaria. Los números en la tabla (a, b, c y d) se refieren al número de pares de observaciones donde los observadores proporcionaron los resultados indicados.

Una medida de la repetibilidad es la concordancia observada (O) o el índice de concordancia de pares medio, que puede calcularse como:

$$(\text{N}^{\circ} \text{ de pares concordantes} / \text{N}^{\circ} \text{ total de pares}) = (a+d)/N$$

Este índice tiene el inconveniente de que incluso si los dos observadores estuvieran clasificando al azar, sería de esperar algo de concordancia entre ellos. El estadístico kappa (κ) es una medida alternativa que tiene en cuenta la concordancia fortuita esperada.

Para calcular el estadístico kappa en primer lugar tiene que calcularse el número de pares de observaciones fortuitas (debidas al azar) que se esperarían en las celdas (++) y (--). El valor esperado en cualquier celda viene dado por

$$[(\text{Total de la fila}) \times (\text{Total de la columna})] / \text{Total final}$$

Así, para la celda (++) el valor esperado será igual a:

$$[(a+b) \times (a+c)] / N$$

y para la celda (--), el valor esperado será igual a:

$$[(c+d) \times (b+d)] / N$$

Ahora podemos calcular la concordancia esperada (E) por azar:

$$[(\text{Valor esperado para la celda } ++) + \text{Valor esperado para la celda } (--)] / N$$

y la concordancia real no fortuita:

Concordancia observada (O) – Concordancia esperada (E)

Este valor es todavía difícil de interpretar ya que pueden obtenerse resultados similares para valores completamente diferentes de O y E . Por ejemplo, la concordancia real no fortuita es igual a 0,20 si $O = 0,95$ y $E = 0,75$ o si $O = 0,75$ y $E = 0,55$. Lo que necesitamos saber es cuanto representa respecto a la concordancia potencial máxima no fortuita que podría haberse alcanzado. La concordancia completa implicaría que todos los resultados estarían en las celdas ($++$) y ($--$), y por tanto, $(a+d)/N$ sería igual a 1. Así, el potencial para la concordancia real no fortuita es

$$1 - E$$

El estadístico kappa indica cuánto representa la concordancia real no fortuita ($O - E$) respecto a este potencial ($1 - E$).

$$\text{Kappa } (\kappa) = (O - E) / (1 - E)$$

El estadístico kappa puede utilizarse de forma similar para determinar la variabilidad intra-observador. Los valores de este coeficiente pueden variar de $-1,0$ a $1,0$. Un valor de $1,0$ indica una concordancia perfecta y un valor de cero significa que no hay una concordancia mejor que la fortuita; un valor negativo indica que el nivel de discordancia es mayor al esperado de forma fortuita. No existe un valor de kappa que pueda considerarse universalmente indicativo de una buena concordancia, pero en la práctica, cualquier valor de κ inferior a $0,5$ indicará una baja concordancia. Landis y Koch (1977) propusieron las siguientes pautas: valores de kappa $\leq 0,4$ representan una concordancia de baja a discreta; $0,41-0,60$ concordancia moderada, $0,61-0,80$ concordancia notable y $0,81-1,00$ concordancia casi perfecta.

El uso del estadístico kappa puede ampliarse a situaciones en las que los resultados del ensayo se clasifican en más de dos categorías como en el Ejemplo 2.5. En este ejemplo, hubo una concordancia sustancial entre los observadores A y B. Se realizaron unos cálculos similares para la concordancia intra-observador. El índice kappa fue igual a $0,83$. En general, la concordancia intra-observador tiende a ser mayor que la concordancia inter-observador.

No obstante, los valores de kappa no deben presentarse solos debido a que proporcionan una medida resumen de concordancia sin ofrecer ninguna indicación sobre donde se habían producido las discordancias. Por tanto, los resultados de un estudio de precisión deben presentarse en tablas similares a la Tabla 2.8, de modo que puedan identificarse las principales áreas de concordancia y discordancia. El estadístico kappa también puede ponderarse si se da diferente importancia a diferentes tipos de concordancia o discordancia (Landis y Koch, 1977).

Ejemplo 2.5. En el estudio de Kiviat *et al.* (1990) mencionado en el Apartado 2.6.1. los autores indican «... para evaluar la variación inter-observador, dos observadores sin conocimiento de otros datos de laboratorio y clínicos revisaron inicialmente todas las autorradiografías y las muestras se clasificaron como positivas, negativas o indeterminadas según las especificaciones del fabricante. La variación inter-observador se evaluó haciendo que el observador A volviera a leer las membranas seis meses después sin conocimiento de otros (o anteriores) datos de laboratorio o clínicos». Los resultados de la variabilidad inter-observador se muestran en la Tabla 2.8.

	Observador B ^b			Total fila
	Positivo	Negativo	Indeterminado	
Observador A ^b				
Positivo	58 (a ₁)	8 (a ₂)	2 (a ₃)	68 (a)
Negativo	12 (b ₁)	357 (b ₂)	3 (b ₃)	372 (b)
Indeterminado	0 (c ₁)	0 (c ₂)	7 (c ₃)	7 (c)
Total por columna	70 (n ₁)	365 (n ₂)	12 (n ₃)	447 (N)

^a Datos de Kiviat *et al.* (1990).

^b Los números en la tabla representan los números de pares de observaciones en los que los observadores proporcionaron los resultados indicados. Las letras entre paréntesis indican cada celda específica en la tabla.

$$\text{Concordancia observada (O)} = (a_1 + b_2 + c_3)/N = (58 + 357 + 7)/447 = 0,94$$

$$\text{Valor esperado para la celda } a_1 = (a \times n_1)/N = (68 \times 70)/447 = 10,65$$

$$\text{Valor esperado para la celda } b_2 = (b \times n_2)/N = (372 \times 365)/447 = 303,76$$

$$\text{Valor esperado para la celda } c_3 = (c \times n_3)/N = (7 \times 12)/447 = 0,19$$

$$\text{Concordancia esperada por azar (E)} = (10,65 + 303,76 + 0,19)/447 = 0,70$$

$$\text{Concordancia real no fortuita (O-E)} = 0,94 - 0,70 = 0,24.$$

$$\text{Concordancia potencial no fortuita} = 1 - 0,70 = 0,30.$$

$$\text{kappa } (\kappa) = 0,24 / 0,30 = 0,80$$

Tabla 2.8. Variabilidad inter-observador en la lectura de la prueba ViraPap^{®a}

Existen métodos similares para evaluar la fiabilidad de las determinaciones que proporcionan resultados en una escala continua (p.e. determinaciones de la presión sanguínea, niveles de glucosa en sangre) en vez de una escala cualitativa. No obstante, estos métodos están más allá del objetivo de este capítulo. Los interesados en ampliar este tema deben referirse a Bland y Altman (1986).

2.7 Consecuencias de los errores de la medición

Los errores en la medición dan lugar de manera invariable a que los individuos no se clasifiquen correctamente y pueden dar lugar a conclusiones falsas sobre la relación entre la exposición y el resultado de interés. El im-

pacto de los errores en la medición sobre los resultados de un estudio epidemiológico depende esencialmente de la naturaleza de la clasificación errónea.

Considere el siguiente ejemplo hipotético. Suponga que queremos determinar si el fumar cigarrillos está asociado con el cáncer de pulmón y que tenemos que confiar en un cuestionario que pregunta '¿ha fumado usted alguna vez?' y '¿tiene usted cáncer de pulmón?'. El cuestionario se administra a 10.000 hombres. Supongamos que la situación tabáquica 'verdadera' de la población de estudio (determinada por un ensayo perfecto, con una sensibilidad y especificidad igual al 100%) es tal y como se muestra en la Tabla 2.9 .

Tabla 2.9. Distribución de una población según hábito tabáquico y estado de la enfermedad tal y como se habría determinado con una prueba perfecta para la medida del hábito tabáquico (sensibilidad y especificidad del 100%). Datos hipotéticos.

		Fumar cigarrillos		
		Alguna vez	Nunca	Total
Cáncer de pulmón	Sí	150	50	200
	No	1.850	7.950	9.800
	Total	2.000	8.000	10.000

La Tabla 2.9 muestra que el cáncer de pulmón es más frecuente entre los que han sido fumadores alguna vez (150 entre 2.000 = 7,5%) que entre los que nunca han sido fumadores (50 de 8.000 = 0,63%). Por tanto, si pudiera utilizarse un instrumento perfecto para medir los hábitos de fumar se observaría que los que han sido fumadores alguna vez tenían una probabilidad doce veces mayor (7,5% / 0,63% = 12) de desarrollar cáncer de pulmón que los que nunca habían sido fumadores.

Tabla 2.10. Distribución observada de una población según hábito tabáquico y estado de la enfermedad mediante cuestionario para medir tabaquismo que tuviera una sensibilidad del 80% y una especificidad del 100%. Datos hipotéticos.

		Fumar cigarrillos		
		Alguna vez	Nunca	Total
Cáncer de pulmón	Sí	$150 - 0.2 \times 150 = 120$	$50 + 0.2 \times 150 = 80$	200
	No	$1.850 - 0.2 \times 1.850 = 1.480$	$7.950 + 0.2 \times 1.850 = 8.320$	9.800
Total		1.600	8.400	10.000

Suponga ahora que cuando aplicamos nuestro cuestionario, un 20% de los fumadores, independientemente del estado de su enfermedad, respondían que no eran fumadores (sensibilidad = 80%) pero que todos los hombres que no eran fumadores respondían que no lo eran (especificidad = 100%). Los resultados obtenidos con este cuestionario imperfecto en el estudio se muestran en la Tabla 2.10 .

Utilizando este cuestionario imperfecto, la proporción de cánceres de pulmón en fumadores es de $120/1.600 = 7,5\%$. Esto es, unas ocho veces la

proporción en los no fumadores ($80/8.400 = 0,95\%$). A pesar de la baja calidad de los datos sobre el fumar proporcionados por el cuestionario, todavía puede percibirse la relación entre el fumar cigarrillos y cáncer de pulmón, aunque no tan fuerte como lo es en realidad.

La clasificación errónea no diferencial se produce cuando la clasificación de la exposición o del resultado final es incorrecta para proporciones iguales de sujetos en los grupos a comparar. En otras palabras, la sensibilidad y especificidad de la medición de la exposición (o resultado) son iguales para ambos, los que tienen la enfermedad y los que no la tienen (o los expuestos y no expuestos). En estas circunstancias, la clasificación errónea es aleatoria (todos los individuos tienen la misma probabilidad de ser clasificados erróneamente).

La clasificación errónea no diferencial da lugar a una clasificación errónea de individuos, reduciendo la confianza que puede depositarse en el resultado de la prueba. Aunque esta variación aleatoria tiene implicaciones importantes en la medicina clínica, en epidemiología, donde el interés principal son grupos en lugar de individuos, es menos preocupante. Ésta es una de las fuerzas de la epidemiología. En el ejemplo anterior, la asociación entre fumar y cáncer de pulmón estaba debilitada por el hecho de que los denominados 'nunca fumadores' eran una mezcla de fumadores y no fumadores. Aunque este tipo de clasificación errónea dificulta el desvelar una asociación entre la exposición y el resultado de interés, esto normalmente puede solventarse aumentando el tamaño de la muestra y/o haciendo determinaciones repetidas (excepto para la clasificación errónea no diferencial de variables de confusión como se comentará en el Capítulo 13). Por lo tanto, el epidemiólogo puede confiar en pruebas simples, baratas y no invasivas que son más apropiadas para estudios en la comunidad a pesar del hecho de que generalmente son menos válidas que las utilizadas en situaciones clínicas.

Éste es un aspecto importante en la investigación epidemiológica que algunos médicos considera difícil de aceptar. El centro de atención de los clínicos son los pacientes individuales de los que intentan obtener la información más completa y válida posible con objeto de proporcionar el mejor diagnóstico y tratamiento posibles. Acostumbrados a utilizar procedimientos especializados y de alta tecnología, es duro para el clínico aceptar que puede realizarse un estudio científico en base al uso de datos de calidad relativamente baja tales como cuestionarios o certificados de defunción.

La clasificación errónea diferencial se produce cuando la sensibilidad y/o especificidad de la determinación de la exposición para el grupo enfermo difiere de la del grupo no enfermó, o cuando la sensibilidad y/o especificidad de la determinación para los expuestos difiere de la del grupo no expuesto. En otras palabras, la clasificación errónea diferencial se produce cuando los errores en la clasificación del resultado dependen del estado de la exposición y viceversa. Por ejemplo, es más probable que las mujeres con cáncer de mama recuerden si han tomado anticonceptivos orales en el pasado que las mujeres sanas. En el ejemplo hipotético considerado anterior-

mente, la clasificación errónea diferencial se habría producido si fuera más probable que los hombres con cáncer de pulmón informaran sobre sus hábitos de fumar con más o menos exactitud que los hombres sin cáncer de pulmón; en estas circunstancias, los datos resultantes podrían haber exagerado o atenuado (o incluso invertido) la relación entre el consumo de tabaco y el cáncer de pulmón.

La clasificación errónea diferencial es una consecuencia de defectos en el diseño o realización de un estudio epidemiológico. Lamentablemente, no puede controlarse en el análisis y su efecto *no* puede reducirse aumentando el tamaño de la muestra.

Una discusión más detallada de las consecuencias de los errores en la determinación de la exposición y el resultado en la interpretación de estudios epidemiológicos se presentará en otros capítulos de este libro, en especial en el Capítulo 13.

2.8 ¿Cómo puede reducirse la clasificación errónea de exposición y resultado?

Todos los procedimientos utilizados en las determinaciones deberían describirse con detalle en el protocolo del estudio, de modo que un segundo grupo de investigadores pudiera, dentro de los límites de variabilidad biológica y física, reproducir las determinaciones obtenidas utilizando sólo esta descripción escrita.

Por lo tanto, en el caso de una entrevista personal, incluirá:

- especificaciones sobre la formación de los entrevistadores e instrucciones que se les proporcionaron,
- instrucciones o explicaciones que los entrevistadores proporcionaron a los sujetos,
- el cuestionario utilizado para obtener los datos de los sujetos,
- los métodos de control de calidad utilizados.

En el caso de una prueba de laboratorio, incluirá:

- métodos para la preparación de los sujetos,
- métodos para la obtención, manipulación, transporte y almacenamiento de las muestras,
- métodos analíticos en el laboratorio,
- métodos de control de calidad.

El epidemiólogo debe mantener un contacto estrecho con los especialistas en el laboratorio de modo que se establezcan al inicio del estudio unos criterios estándar para la toma de muestras, almacenamiento y análisis. En general, la mayoría de los laboratorios aplican de manera rutinaria métodos de control de calidad intra e inter-laboratorio. A pesar de esto, es aconsejable remitir las muestras de los estudios epidemiológicos sin indicar el estado de la exposición de los sujetos de los cuales se obtuvieron. En ocasiones se remiten réplicas de muestras sin que el analista sepa que son réplicas.

Los métodos de determinación siempre deben evaluarse previamente en

un estudio piloto para identificar problemas potenciales, para proporcionar indicaciones de su validez y fiabilidad, y evaluar de qué modo los observadores o las personas entrevistadas pueden estar sesgados. Estos aspectos se seguirán comentando en otros capítulos, en especial en los Capítulos 13 y 18.

2.9 Fuentes de datos recogidos sistemáticamente

Los 'datos de rutina' son datos derivados de sistemas continuos de obtención de datos asociados a los servicios sociales y sanitarios. En general, estos datos no se obtienen específicamente para responder a una pregunta en particular. Los datos se almacenan en algún tipo de índice o base de datos que puede utilizarse con fines epidemiológicos, se obtuvieran o no con este fin. Éstos incluyen datos de censos y registros de población, certificados de defunción, registros de cáncer, certificados de nacimiento, sistemas de información sanitaria, registros médicos, registros hospitalarios, etc. (véase Capítulo 11).

Estos sistemas continuos de obtención de datos pueden proporcionar información sobre la(s) exposición(es) y resultados(s) de interés en un estudio epidemiológico. Dos de estos sistemas tienen una importancia especial en epidemiología del cáncer – los certificados de defunción y los registros de cáncer.

2.9.1 Certificado de defunción

Los datos de mortalidad normalmente se basan en un certificado de defunción estándar que registra la fecha de la muerte, causa de muerte, edad, sexo, fecha de nacimiento y lugar de residencia. Además, puede registrarse la ocupación y otras variables. En la mayoría de los países, los certificados de defunción los suele cumplimentar un médico u otro trabajador sanitario, pero en algunos casos los puede realizar la policía u otras autoridades. Una vez cumplimentados estos certificados, la causa de muerte se codifica según la Clasificación Internacional de Enfermedades, de la que actualmente se está iniciando la décima revisión (OMS, 1992). Se trata de una clasificación jerárquica que va desde categorías amplias a una clasificación detallada de 4 dígitos (véase Anexo 2.2). Normalmente, sólo se codifica la causa de muerte fundamental la cual se utiliza en las estadísticas de mortalidad, pero también pueden codificarse las causas contribuyentes.

Aunque son más completos y precisos que muchas fuentes de rutina de datos, los datos de mortalidad todavía adolecen de numerosos errores de clasificación (Cameron y McGoogan, 1981; Heasman y Lipworth, 1966). Un estudio internacional extenso realizado en varias ciudades de Inglaterra, EE.UU. y Latinoamérica que incluyó 8.737 muertes por cáncer, mostró que un 20% de las muertes clasificadas como cáncer en el certificado de defunción era incorrecto (en realidad eran debidas a otras causas) (Puffer y Griffith, 1967). No obstante, un 24,6% de los casos reales de cáncer se habían clasificado erróneamente como otras causas de muerte. Por lo tanto, globalmente, la mortalidad total por cáncer sólo se infraestimó en un 4%

en las estadísticas oficiales derivadas de los certificados de defunción. La proporción de errores en la clasificación varió según el tipo de cáncer, siendo mayor en los cánceres más difíciles de diagnosticar (p.e. cáncer de hígado primario, tumores cerebrales).

Las estadísticas internacionales de mortalidad por cáncer se publican regularmente por la Organización Mundial de la Salud (*World Health Statistics Annual Series*) y por Segi y sus colaboradores (Kurihara *et al.*, 1989).

2.9.2 Registros de cáncer

Existen dos tipos de registros de cáncer: los hospitalarios y los poblacionales. Los registros de cáncer hospitalarios se ocupan de la anotación de todos los pacientes con cáncer observados en un hospital en particular. El principal objetivo de estos registros es contribuir al cuidado del paciente y al tratamiento administrativo, aunque pueden utilizarse, en cierto grado, con fines epidemiológicos. Por ejemplo, pueden realizarse estudios caso-control para investigar la etiología de un determinado cáncer comparando las características de estos casos con las de un grupo control; este último grupo puede estar formado por pacientes con otros tipos de cáncer o por otros pacientes del hospital. No obstante, estos registros no pueden proporcionar medidas de la aparición de cáncer en la población general debido a que no es posible definir la población de la que proceden los casos.

Los registros de cáncer poblacionales pretenden registrar todos los casos de cáncer nuevos (incidentes) en una población bien definida. Como resultado, y a diferencia de los registros hospitalarios, proporcionan medidas de la aparición de cáncer en la población de origen. El registro de cáncer poblacional se ha desarrollado en muchos países para proporcionar unos datos razonablemente comparables sobre la incidencia del cáncer y como recurso para estudios epidemiológicos. La Agencia Internacional de Investigación sobre el Cáncer ha compilado los datos de incidencia del cáncer de los registros de mejor calidad en volúmenes sucesivos de la serie *Incidencia del Cáncer en Cinco Continentes* [*Cancer Incidence in Five Continents*] (Doll *et al.*, 1966; Waterhouse *et al.*, (1970, 1976, 1982); Muir *et al.*, 1987; Parkin *et al.*, 1992, 1997). En las tablas de los diferentes volúmenes se presentan algunos indicadores de la calidad de los datos según los diferentes registros incluidos en esta publicación. No obstante, la mayoría de las medidas presentadas son sólo indicadores indirectos de calidad de los datos: proporción de registros comprobados histológicamente, proporción de casos registrados sólo en base a la información del certificado de defunción; proporción de casos en los que falta información, etc. Existen análisis más sistemáticos sobre la validez de estos datos en algunos registros en los que una muestra de casos se volvió a tomar y a recodificar (Parkin *et al.*, 1994).

La mayoría de los registros de cáncer reúnen información adicional sobre los pacientes con cáncer como ocupación, clase social, país de nacimiento, raza, etc. Por lo tanto, la aparición de cáncer puede evaluarse además según estas variables.

El papel de los registros de cáncer en la epidemiología del cáncer se comentará en detalle en el Capítulo 17.

2.9.3 Enlace de registros

En los registros de muchas instituciones y agencias existe información sobre las características de individuos desde el nacimiento a la muerte. Estos registros pueden unirse en un registro único completo utilizando identificadores personales (enlace de registros). Un registro así unificado puede utilizarse para investigaciones epidemiológicas y de salud pública. La posibilidad de relación entre los registros varía enormemente entre países en función de cómo se haya obtenido e identificado la información. En los países nórdicos, a cada persona se le asigna un número que se utiliza como identificador en la seguridad social, censo y registros sanitarios. En este contexto, los datos de mortalidad e incidencia del cáncer pueden obtenerse con bastante facilidad y relacionarlos con otras series de datos de interés. En el Reino Unido existe un registro nacional relacionado con el servicio sanitario que se utiliza ampliamente para estudios de seguimiento de cáncer y mortalidad. El registro está informatizado y permite el seguimiento de las personas que estaban vivas en enero de 1991, identificadas por su nombre y fecha de nacimiento.

El enlace de los archivos de los registros de cáncer con los archivos de otras fuentes, tales como los datos del censo y registros de industrias, ha sido utilizado para investigar factores de riesgo para cánceres ocupacionales y cánceres relacionados con la reproducción. Los registros hospitalarios también pueden proporcionar información sobre exposición ya que con frecuencia incluyen los números de historia clínica. Este enlace con los registros hospitalarios se ha utilizado en estudios sobre riesgos de cáncer asociados a radioterapia y otros tratamientos (Day y Boice, 1983; Kaldor *et al.*, 1992).

El Estudio de Enlace de Registros de Oxford (ORLS) realizado en 1962 en Oxford (Inglaterra) y el sistema nacional escocés de enlace de registros médicos proporcionaron dos buenos ejemplos de este tipo de estudios. El ORLS se realizó para evaluar la factibilidad, coste y métodos para enlazar registros sanitarios en una comunidad. Este sistema relaciona información de morbilidad y mortalidad y proporciona información sobre una amplia serie de variables. Es posible utilizar sus datos para estudiar una variedad de cuestiones etiológicas y evaluar la historia natural de varias enfermedades (Acheson, 1967; Baldwin *et al.*, 1987). En Escocia, pueden relacionarse todos los registros de nacimientos, muertes, hospitalizaciones, incidencia de cáncer, exámenes médicos escolares y de niños minusválidos (Heasman y Clarke, 1979). Se han realizado sistemas similares de relación de registros en los Estados Unidos, el *National Center for Health Statistics* (Feinleib, 1984), así como en muchos otros países desarrollados.

Lecturas adicionales

* Armstrong *et al.* (1992) proporciona una cobertura completa de los principios y aspectos prácticos de diseño de cuestionarios, realización de entrevistas personales, extracción de datos de registros y uso de medidas biológicas y medioambientales. Aunque este libro se centra en la medición de la exposición, muchos de los principios presentados también son importantes para la medida de resultados.

* Un artículo referenciado con frecuencia sobre la validez y fiabilidad de las pruebas que proporcionan resultados sobre una escala continua (p.e. mediciones de la presión sanguínea) es el de Bland y Altman (1986).

* Para más discusión (y más compleja) sobre el estadístico kappa véase Feinstein y Cicchetti (1990), Cicchetti y Feinstein (1990) y Lantz y Nebenzahl (1996).

Cuadro 2.1. Puntos esenciales

- En estudios epidemiológicos sobre el cáncer es necesario determinar: 1) la(s) exposición(es) primaria(s) de interés; 2) otra(s) exposición(es) que puedan influenciar el resultado (variables potenciales de confusión) y 3) el propio(s) resultado(s).
- Existe una gran variedad de métodos para determinar la exposición y la enfermedad. Estos incluyen entrevistas personales, cuestionarios auto-administrados, diarios, registros, medidas biológicas y mediciones en el medio ambiente. Cada método tiene sus propias ventajas e inconvenientes.
- En todo estudio epidemiológico es importante evaluar la validez y fiabilidad de las principales determinaciones de la exposición y resultado. Éstas proporcionarán una estimación de la magnitud de los errores de medición y su posible impacto sobre los resultados del estudio. Los errores de medida pueden ser diferenciales y no diferenciales.
- El *error de medida no diferencial* se produce cuando la sensibilidad y especificidad de la medición de la exposición para el grupo con enfermedad son iguales que para el grupo sin la enfermedad (o cuando la sensibilidad y especificidad de la determinación de la enfermedad son iguales para los sujetos expuestos y no expuestos). Este tipo de error de medición generalmente da lugar a una infraestimación de la asociación entre la exposición y el resultado de interés. Aunque el error no diferencial dificulta el desvelar una asociación entre la exposición y el resultado de interés, normalmente puede resolverse aumentando el tamaño de la muestra y/o repitiendo las determinaciones.
- El *error de medida diferencial* se produce cuando la sensibilidad y/o especificidad de la determinación de la exposición para el grupo de sujetos enfermos difiere de la de los sujetos no enfermos (o cuando la sensibilidad y/o especificidad de la determinación de la enfermedad es diferente para los sujetos expuestos y no expuestos). Este tipo de error puede exagerar o atenuar (o incluso invertir) la relación entre la exposición y el resultado de interés de modo que los resultados del estudio pueden ser totalmente erróneos. Lamentablemente, esto no puede controlarse en el análisis y su efecto no puede reducirse aumentando el tamaño de la muestra.

Anexo 2.1

Diseño de un cuestionario

Los cuestionarios se utilizan en epidemiología para evaluar los niveles de exposición a posibles agentes causales y más raramente para determinar la presencia o ausencia de enfermedad o cualquier otro resultado de interés.

A.2.1.1 Objetivos del diseño de un cuestionario

- (1) Proporcionar medidas válidas de la(s) exposición(es) y el resultado(s) en estudio.
- (2) Ser cumplimentado fácilmente por el entrevistador y/o el sujeto.
- (3) Facilitar el procesamiento y análisis de los datos.

A.2.1.2 Principios generales del diseño de un cuestionario

Contenido

El cuestionario debe ser lo más breve posible; cada pregunta debe estar justificada en relación con los objetivos del estudio. Es importante asegurarse de que las variables necesarias para el análisis se puedan obtener fácilmente del cuestionario.

Tipos de preguntas

Existen dos tipos principales de preguntas: 'abiertas' y 'cerradas'.

Las *preguntas abiertas* permiten a los entrevistados responder en sus propios términos y deben registrarse con las propias palabras del entrevistado. Las preguntas abiertas deben utilizarse para datos numéricos (por ejemplo, edad, fecha de nacimiento) y para preguntas para las cuales hay muchas respuestas posibles (p.e. país de nacimiento).

Ejemplo A.2.1.1. Ejemplo de una pregunta abierta.

¿Cuál es su lengua materna?.....

Las *preguntas cerradas* permiten sólo una serie limitada de respuestas. El cuestionario debe especificar con detalle todas las posibles respuestas alternativas. Con respuestas alternativas múltiples debe proporcionarse una alternativa final 'Otros, por favor especificar...' a no ser que se hayan incluido todas las respuestas posibles. También debe incluirse una opción 'No sabe' para preguntas en las que es posible que algunos sujetos no sepan

(o no recuerden) la respuesta. Debe incluirse una opción 'No procede' si la pregunta no procede para todos los sujetos.

Ejemplo A.2.1.2. Ejemplo de una pregunta cerrada.

7. Si nunca ha estado embarazada fue porque:	
Nunca lo intentó	<input type="checkbox"/> 1
Lo intentó pero no ocurrió	<input type="checkbox"/> 2
Otras razones: especificar por favor	<input type="checkbox"/>
No procede	<input type="checkbox"/> 7

Los cuestionarios epidemiológicos suelen contener una mayoría de preguntas cerradas para reducir la posibilidad de sesgos debidos al entrevistador, a la respuesta, a la interpretación y/o a la codificación, y para facilitar el procesamiento de los datos.

Redacción de las preguntas

Las preguntas deben escribirse en un lenguaje simple, no amenazador, evitando el uso de abreviaciones y jerga técnica. La redacción debe evitar cualquier indicación sobre la preferencia de una respuesta en particular por parte de los investigadores. Cada pregunta debe contener sólo un concepto relacionado con un periodo de tiempo claro.

Orden de las preguntas

Las preguntas deben seguir una secuencia lógica que se parezca lo más posible a la secuencia que los entrevistados puedan esperar para seguir los pensamientos sobre el tema. Las preguntas sobre un tema en particular deben agruparse y realizarse de lo general a lo particular. Cuando una respuesta a una pregunta general hace que otras respuestas sobre ese tema sean irrelevantes (p.e. una mujer que nunca ha estado embarazada no tiene que responder preguntas sobre el número y las características de los embarazos), puede introducirse una ramificación de la secuencia de preguntas. Este mecanismo debe ser lo más simple posible con instrucciones claras indicadas en el cuestionario (Ejemplo A2.1.3).

Presentación del cuestionario

La presentación de un cuestionario es importante tanto en los cuestionarios auto-administrados como en los administrados por un entrevistador. Una presentación agradable incitará interés y animará a su cumplimentación correcta. Una página separada con una breve introducción,

Ejemplo A.2.1.3. Ejemplo de instrucciones para omitir preguntas (ramificación).

1. ¿Ha estado embarazada alguna vez?

Sí 1

No 2

Si es NO, por favor vaya a la pregunta 4.

Si es SÍ,

2. ¿Cuántos embarazos ha tenido en total?

(incluyendo nacidos muertos, pérdidas y abortos).

notas explicatorias e instrucciones debe preceder a la primera pregunta. Los cuestionarios largos pueden subdividirse en apartados, cada uno correspondiente a un tema específico, para ayudar a los entrevistadores y a los sujetos a completarlos. A todas las preguntas se les debe asignar un número.

Si algunas preguntas son opcionales, debe indicarse en el cuestionario con instrucciones claras y explicaciones de salto y ramificación adecuadas. Para las preguntas que se repiten varias veces tales como las preguntas sobre cada embarazo, puede utilizarse una presentación en forma de tabla (Ejemplo A2.1.4).

Al final del cuestionario debe dejarse espacio para cualquier información o comentarios que el sujeto pueda querer hacer.

Método de administración

El cuestionario puede ser auto-administrado o administrado por un entrevistador. En general, los cuestionarios auto-administrados deben ser más simples y diseñados con mucha más precaución que los destinados a ser utilizados por entrevistadores.

Registro y codificación de respuestas

La mayoría de los cuestionarios se prepararán para permitir la codificación numérica de todas las respuestas para su procesamiento por el ordenador. A cada respuesta posible en el formulario se le asigna un código (como en los Ejemplos A2.1.2 a A2.1.4). Los datos numéricos (p.e. número de embarazos) no requieren la codificación ya que puede introducirse el número exacto. Pero incluso con estos cuestionarios «precodificados» todavía será necesaria la codificación después de la obtención de los datos para algunas

Ejemplo A.2.1.4. Ejemplo de pregunta con presentación en forma de tabla.

3. Por favor indique las características de sus embarazos:					
	1º	2º	3º	4º	5º
Edad al inicio del embarazo (años)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Resultado	Nacidos vivos <input type="checkbox"/> 1 Nacidos muertos <input type="checkbox"/> 2 Pérdida <input type="checkbox"/> 3 Abortos <input type="checkbox"/> 4	Nacidos vivos <input type="checkbox"/> 1 Nacidos muertos <input type="checkbox"/> 2 Pérdida <input type="checkbox"/> 3 Abortos <input type="checkbox"/> 4	Nacidos vivos <input type="checkbox"/> 1 Nacidos muertos <input type="checkbox"/> 2 Pérdida <input type="checkbox"/> 3 Abortos <input type="checkbox"/> 4	Nacidos vivos <input type="checkbox"/> 1 Nacidos muertos <input type="checkbox"/> 2 Pérdida <input type="checkbox"/> 3 Abortos <input type="checkbox"/> 4	Nacidos vivos <input type="checkbox"/> 1 Nacidos muertos <input type="checkbox"/> 2 Pérdida <input type="checkbox"/> 3 Abortos <input type="checkbox"/> 4
Duración del embarazo (semanas)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Lactancia	Sí <input type="checkbox"/> 1 No <input type="checkbox"/> 2 No procede <input type="checkbox"/> 7 No sabe <input type="checkbox"/> 9	Sí <input type="checkbox"/> 1 No <input type="checkbox"/> 2 No procede <input type="checkbox"/> 7 No sabe <input type="checkbox"/> 9	Sí <input type="checkbox"/> 1 No <input type="checkbox"/> 2 No procede <input type="checkbox"/> 7 No sabe <input type="checkbox"/> 9	Sí <input type="checkbox"/> 1 No <input type="checkbox"/> 2 No procede <input type="checkbox"/> 7 No sabe <input type="checkbox"/> 9	Sí <input type="checkbox"/> 1 No <input type="checkbox"/> 2 No procede <input type="checkbox"/> 7 No sabe <input type="checkbox"/> 9

preguntas abiertas o para la categoría «Otros. Por favor, especificar» de las preguntas cerradas (como en el Ejemplo A2.1.2).

La codificación de los cuestionarios puede ser una tarea compleja y puede ser necesario desarrollar un manual de codificación con normas específicas de codificación. Se han desarrollado y publicado diferentes sistemas de clasificación que pueden utilizarse para codificar cánceres por su topografía (p.e. *Clasificación Internacional de Enfermedades* (OMS, 1992)) y por su morfología y comportamiento (p.e. *CIE-O*, (Perey *et al.*, 1990)) (véase Anexo 2.2), ocupaciones (p.e. *Clasificación de ocupaciones* (OPCS, 1970)) y muchas otras variables.

A.1.2.3 Evaluación de un cuestionario

Los cuestionarios deben someterse a dos formas de evaluación:

Ensayo previo

Todos los cuestionarios deben someterse a un ensayo previo. Este incluye el ensayo de borradores del cuestionario con muestras de sujetos similares a los que se pretende estudiar. Su objetivo es identificar preguntas que se entienden mal, son ambiguas o incitan respuestas hostiles o indeseables. Los ensayos previos del cuestionario deben efectuarse tal y como se administrará finalmente. Se debe pedir a los entrevistadores y a los sujetos del estudio que participen y las preguntas deben revisarse en base a los comentarios. Normalmente serán necesarios múltiples ensayos previos antes de conseguir la forma final de un cuestionario.

Evaluación de la validez

La validez de un cuestionario como medida de las variables sobre las que se busca información siempre debe determinarse en una muestra de sujetos antes del inicio del estudio principal. Esto requiere una comparación de los resultados obtenidos utilizando el cuestionario con los obtenidos utilizando un 'estándar de oro' (véase Apartado 2.6). Por ejemplo, las preguntas sobre hospitalizaciones en el pasado e intervenciones quirúrgicas pueden validarse frente a los registros hospitalarios. La validación normalmente es difícil, con frecuencia cara, y algunas veces imposible debido a la falta de un 'estándar de oro' apropiado.

A.2.1.4 Utilización de cuestionarios estándar

Cuando existe un cuestionario estándar para la medida de una exposición particular, puede ser deseable utilizarlo para ahorrar el tiempo y el esfuerzo implicados en el diseño de un cuestionario nuevo. Además, el cuestionario estándar se habrá utilizado ampliamente y habrá resultado satisfactorio e incluso puede haber sido validado (aunque la validez en una población puede no asegurar la validez en otra). También permitirá la comparación de sus datos con los obtenidos por otros investigadores.

Puede ser necesario hacer algunos cambios en el formato del cuestionario estándar para adecuarlo a la población en estudio. Estos cambios pueden afectar a la validez del cuestionario. La validez de una modificación puede ser evaluada a su vez frente al cuestionario original.

En Armstrong *et al.* (1992) se presenta una discusión completa sobre diseño del cuestionarios.

Anexo 2.2.

Clasificación de enfermedades

Los neoplasias pueden clasificarse de muchas maneras pero las clasificaciones más importantes para los epidemiólogos son las basadas en:

- (1) Topografía, es decir, el órgano donde se origina el tumor.
- (2) Morfología (o histología), es decir, el aspecto microscópico del tumor.
- (3) Comportamiento, es decir, la tendencia a invadir otros tejidos (maligno, benigno, *in situ* e incierto)

Las definiciones uniformes y los sistemas uniformes de clasificación de enfermedades son fundamentales para el estudio cuantitativo de enfermedades. Sin una herramienta estándar de clasificación que permanezca fija durante periodos de tiempo y que se aplique uniformemente de un lugar a otro, no sería posible realizar análisis comparativos significativos de los datos de morbilidad y mortalidad. La Clasificación Internacional de Enfermedades (CIE) [*International Classification of Diseases (ICD)*] publicada por la Organización Mundial de la Salud es esta herramienta estándar. Esta clasificación se revisa cada diez años aproximadamente (Tabla A.2.2.1); la décima revisión (CIE-10) (OMS, 1992) es la actualmente vigente. En la introducción de la CIE-7 (OMS, 1957) se incluye una revisión histórica de la clasificación de enfermedades desde su primera aparición, la Clasificación Bertillon de Causas de Muerte, hasta 1947. La revisión de la clasificación de 1948 a 1985 aparece en Muir y Percy (1991).

Revisión	Año de publicación	Organización
1ª (CIE-1)	1900	Gobierno francés
2ª (CIE-2)	1910	
3ª (CIE-3)	1920	
4ª (CIE-4)	1929	Organización Sanitaria de la Liga de Naciones
5ª (CIE-5)	1938	
6ª (CIE-6)	1948	Organización Mundial de la Salud
7ª (CIE-7)	1957	
8ª (CIE-8)	1967	
9ª (CIE-9)	1977	
10ª (CIE-10)	1992	

Tabla A2.2.1.
Manuales de la *Clasificación Internacional de Enfermedades*

Aunque manteniendo la estructura tradicional de la CIE-9, la décima revisión de la CIE utiliza un esquema de codificación alfanumérico (es decir, el primer carácter de la categoría es una letra) que sustituye a los códigos

numéricos de la CIE-9 y revisiones anteriores. Esto permite intervalos de codificación más amplios y deja margen para futuras inclusiones de nuevas enfermedades sin alterar el sistema de numeración.

La CIE-10 tiene tres volúmenes. El volumen 1 incluye la lista tabulada de la clasificación al nivel de tres y cuatro caracteres, tabulaciones especiales de morbilidad y mortalidad, definiciones y regulaciones de nomenclatura. El volumen 2 es esencialmente un manual de instrucciones y el volumen 3 proporciona un índice alfabético.

El capítulo en los manuales de CIE que trata de neoplasias es principalmente una clasificación topográfica ordenada según la localización anatómica del tumor, a excepción de unos pocos tipos histológicos tales como linfomas y leucemias (Tabla A2.2.2). Los órganos están ordenados según los sistemas. Las neoplasias con un comportamiento determinado se agrupan en bloques: malignas, benignas, *in situ* y de comportamiento indeterminado.

Tabla A2.2.2.
Clasificación de neoplasias según la
CIE-10 (OMS, 1992)

C00-C75	Neoplasias malignas, confirmadas o supuestamente primarias, de lugares específicos, excepto de tejido linfático, hematopoyético y tejidos relacionados.
C00-C14	Labio, cavidad oral y faringe
C15-C26	Órganos digestivos
C30-C39	Órganos respiratorios e intratorácicos
C40-C41	Hueso y cartilago articular
C43-C44	Piel
C45-C49	Mesotelio y tejidos blandos
C50	Mama
C51-C58	Órganos genitales femeninos
C60-C63	Órganos genitales masculinos
C64-C68	Tracto urinario
C69-C72	Ojos, cerebro y otras partes del sistema nervioso central
C73-C75	Tiroides y otras glándulas endocrinas
C76-C80	Neoplasias malignas mal definidas, secundarias y lugares inespecíficos
C81-C96	Neoplasias malignas, confirmadas o supuestamente primarias, de tejido linfático, hematopoyético y tejidos relacionados.
C97	Neoplasias malignas de múltiples lugares independientes (primarios)
D00-D09	Neoplasias <i>in situ</i>
D10-D36	Neoplasias benignas
D37-D48	Neoplasias de comportamiento no claro o desconocido

La primera clasificación morfológica se desarrolló en 1951 y desde entonces han aparecido otras (Tabla A.2.2.3). Las más utilizadas han sido la MOTNAC (Manual de Nomenclatura de Tumores y de Codificación) (American Cancer Society, 1951; Percy *et al.*, 1968) y, más recientemente, la Clasificación Internacional de Enfermedades para Oncología (CIE-O) (OMS, 1970; Percy *et al.*, 1990). Éstas proporcionan no sólo códigos de morfología y comportamiento, sino también códigos topográficos relacionados directamente con los códigos CIE.

Año de publicación	Manual de código morfológico	Organización	Características principales
1951	<i>Manual de Nomenclatura de Tumores y Codificación (MOTNAC) 1ª edición</i>	Sociedad Americana del Cáncer	Códigos de morfología Códigos de comportamiento
1956	<i>Código Estadístico para Tumores Humanos (STAT CODE)</i>	Organización Mundial de la Salud	Códigos de topografía de CIE-7 Códigos de morfología de MOTNAC Códigos de comportamiento de MOTNAC
1965	<i>Nomenclatura Sistemática de Patología (SNOP) (Apartados 8,9 – Neoplasias)</i>	Colegio de Anatomopatólogos Americanos	Códigos de topografía no relacionados a la CIE Códigos de morfología
1968	<i>Manual de Nomenclatura de Tumores y Codificación (MOTNAC) 2ª edición</i>	Sociedad Americana del Cáncer	Códigos de topografía de CIE-9 Códigos de morfología de SNOP
1976	<i>CIE-O, 1ª edición</i>	Organización Mundial de la Salud	Códigos de topografía de CIE-9 Códigos de morfología de MOTNAC (con extensión de 1 dígito) Códigos de comportamiento de MOTNAC
1977	<i>Nomenclatura Sistemática de Medicina (SNOMED) (Apartados 8,9 – Neoplasias)</i>	Colegio de Anatomopatólogos Americanos	Revisión de SNOP Códigos de topografía no relacionados con CIE Códigos de morfología de CIE-O
1990	<i>CIE-O, 2ª edición</i>	Organización Mundial de la Salud	Códigos de topografía de CIE-10 Códigos de morfología de CIE-O, 1ª edición Códigos de comportamiento de CIE-O, 1ª edición

Tabla A.2.2.3.
Clasificaciones de neoplasias según la morfología y el comportamiento

Una discusión completa de las ventajas e inconvenientes de cada una de estas clasificaciones se presenta en Muir y Percy, (1991). La principal ventaja de la CIE es que es realmente internacional, siendo utilizada por todos los Estados Miembros de la OMS para las tabulaciones de causas de muerte y para la mayoría de las estadísticas sanitarias. El principal inconveniente es que, para la mayoría de localizaciones, no es posible una separación morfológica en la CIE. Por consiguiente, se recomienda generalmente que las agencias interesadas en la identificación de la localización y morfología de los tumores, como los registros de cáncer y laboratorios de patología, utilicen la CIE-O, que es una clasificación de doble eje que proporciona sistemas de codificación independientes para la topografía y la morfología.

Debido a que se están utilizando nuevas clasificaciones y nuevas revisiones de la CIE y CIE-O, los registros de cáncer tienen que convertir los datos codificados por clasificaciones anteriores a los códigos nuevos, con objeto de notificar series a largo plazo. El Instituto Nacional del Cáncer en los EE.UU. ha producido una serie de tablas de conversión para neoplasias (p.e., Percy, 1980, 1981, 1983; Percy y van Holten, 1979). También se han publicado tablas resumen de equivalencia entre varias revisiones de la CIE en algunos volúmenes de Incidencia del Cáncer en Cinco Continentes [*Cancer Incidence in Five Continents*] (p.e., Waterhouse *et al.*, 1976; Muir *et al.*, 1987). La Agencia Internacional de Investigación sobre el Cáncer (IARC) ha desarrollado programas informáticos para efectuar conversiones de CIE-O (primera edición) a CIE-O (segunda edición) y vice-versa, de CIE-O (primera y segunda ediciones) a CIE-9, y de CIE-O (segunda edición) a CIE-10 y están disponibles en disquete para su uso en microordenadores (Ferlay, 1994).

Capítulo 3

Descripción y presentación de datos

Todos los estudios epidemiológicos implican la obtención de datos de las exposiciones y de los resultados de interés. En un estudio bien planeado, las observaciones originales que constituyen los datos contienen la información que satisface los objetivos del estudio. El objetivo en el análisis de datos es extraer de manera concisa la información pertinente a partir de estas observaciones originales.

La primera etapa consiste en la preparación de un formulario apropiado para obtener y codificar los datos relevantes en un formato que esté listo para su introducción en un fichero de ordenador; esta etapa se denomina procesamiento de datos. La segunda etapa consiste en una revisión de los datos registrados, controlando la exactitud, consistencia y totalidad; este proceso se denomina edición de datos. A continuación, el investigador debe resumir los datos en una forma concisa para el análisis subsiguiente; esto generalmente se efectúa presentando la distribución de las observaciones según unas características clave en tablas, gráficos y medidas básicas de resumen. Esta etapa del análisis se denomina reducción de datos. Únicamente después del procesamiento, edición y reducción de datos, el investigador debe realizar una manipulación estadística más elaborada de los datos.

3.1 Procesamiento de datos

Todas las diferentes etapas del procesamiento de datos deben planificarse en la etapa de diseño del estudio, antes del inicio de cualquier obtención de datos. Todos los formularios a utilizar para el registro de los datos obtenidos en un estudio deben diseñarse atentamente y probarse previamente para garantizar que los datos puedan extraerse fácilmente para su procesamiento. Este principio general es aplicable a todas las formas de obtención de datos: cuestionarios, así como formularios para el registro de resultados de pruebas de laboratorio, formularios de extracción de datos de historias clínicas, etc.

La mayoría de los estudios epidemiológicos implican la obtención de grandes cantidades de datos que no son fáciles de manipular. Afortunadamente, los microordenadores están disponibles actualmente a precios razonablemente accesibles (véase Capítulo 18). Para poder introducir los datos en un fichero de ordenador es necesario codificarlos antes. Por ejemplo, el sexo puede codificarse 1 para hombres y 2 para mujeres; en el fichero del ordenador sólo se introducirá el número 1 ó 2. Los datos numéricos (p.e. número de niños) no requieren una codificación ya que puede

introducirse el número exacto. La mayoría de los formularios de obtención de datos están diseñados de modo que a cada respuesta posible se le asigne un código. Pero, incluso con estos formularios «precodificados», será necesaria la codificación después de la obtención de los datos para las «preguntas abiertas» o para la categoría «otros» de las «preguntas cerradas». En el Anexo 2.1 y el Apartado 18.3.6 se presenta una discusión más detallada de estos temas.

Para cada tipo de formulario, se creará un fichero de ordenador para introducir los datos. Es posible diseñar un programa de entrada de datos de modo que la pantalla del ordenador se parezca a la estructura del formulario. Esto ayudará a reducir los errores de la entrada de datos realizados por el personal encargado de la introducción de los mismos.

3.2 Edición de datos

La edición básica de los datos implica la comprobación de cada variable en cuanto a valores ilógicos o inusuales. Por ejemplo, el sexo puede codificarse 1 para hombres y 2 para mujeres. Normalmente se utiliza un valor aparte, puede que 9, para indicar un valor desconocido. Es preferible asignar códigos específicos a una información desconocida o ausente que dejarlos en blanco porque es difícil saber si un campo en blanco corresponde a datos que realmente faltan o a datos que por error no se han introducido. No obstante, debería evitarse el código cero debido a que algunos ordenadores o programas pueden interpretar la información que falta como cero. El intervalo y distribución de cada variable debe examinarse y los valores inadmisibles deben comprobarse comparándolos con los formularios de datos originales.

Además de comprobar los valores incorrectos o inusuales, debe examinarse la distribución de cada variable para ver si es razonable. Esta evaluación puede mostrar problemas importantes que de otro modo podrían no identificarse. También es necesario realizar una comprobación cruzada de la compatibilidad de los códigos para las variables relacionadas. Por ejemplo, los hombres no se deben haber hospitalizado por histerectomía o las mujeres por prostatectomía. La edición atenta implicará muchos de estos controles de compatibilidad que pueden realizarse mediante programas informáticos diseñados para este fin.

La mayoría de los programas de introducción de datos pueden editar los datos de manera interactiva a medida que se introducen. Los valores inadmisibles pueden rechazarse y corregir inmediatamente programando el ordenador para que imprima un mensaje de error en la pantalla y emita una señal de alerta audible para avisar del error a la persona que los introduce. Un programa sofisticado de introducción de datos también puede comprobar la compatibilidad entre variables y puede eliminar algunas incompatibilidades potenciales proporcionando automáticamente códigos apropiados. Por ejemplo, si el sujeto es un hombre, el programa puede proporcionar automáticamente el código adecuado para «ha estado embarazada alguna vez» y saltarse la pregunta. Pero incluso con la edición más sofisticada, durante la

entrada de los datos, sigue siendo importante editarlos antes del análisis para comprobar su totalidad y la racionalidad de la distribución de cada variable. Los programas de entrada de datos no pueden evaluar estos aspectos (véase Apartado 18.3.6).

Se recomiendan tres precauciones básicas para reducir al mínimo errores que aparecen durante la manipulación de los datos. La primera es evitar la copia innecesaria de los datos de un formulario a otro. La segunda es utilizar un procedimiento de comprobación durante la introducción de los datos. Los datos siempre deben introducirse dos veces, preferentemente por dos personas distintas y comparar posteriormente las dos series de datos para resolver las posibles incompatibilidades. La tercera es comprobar todos los cálculos atentamente, ya sea repitiéndolos o, por ejemplo, comprobando que la adición de los subtotales proporciona el global total correcto. Cuando se utiliza un ordenador todos los procedimientos deben probarse inicialmente con una pequeña muestra de datos y los resultados deben comprobarse a mano.

3.3 Reducción de los datos

Después de editar los datos, el investigador debe examinarlos utilizando tabulaciones simples, gráficos y medidas básicas de resumen. La elección de los métodos correctos para presentar y resumir los datos dependerá de la información obtenida, ya que los diferentes tipos de datos deben presentarse y resumirse de diferentes maneras. En el resto de este capítulo consideraremos formas de presentar y resumir dos tipos principales de datos: variables cuantitativas y variables categóricas (o cualitativas).

3.3.1 Datos cuantitativos

Las variables cuantitativas pueden tomar un valor numérico en una escala continua (p.e. edad, peso, altura) o pueden ser números enteros que representan recuentos de algún acontecimiento en particular (p.e. número de niños, número de parejas sexuales).

Presentación de datos cuantitativos

La frecuencia con que aparecen los diferentes valores posibles de una variable en un grupo de sujetos se denomina distribución de frecuencia de la variable. Por ejemplo, podemos querer presentar la distribución de la altura (en cm) de una muestra de 1.250 mujeres examinadas en una clínica para cribaje de cáncer de mama. Debido a que la altura se mide en una escala continua, puede tomar muchos valores distintos y por tanto es más útil agrupar primero los valores antes de presentar la distribución (Tabla 3.1).

El porcentaje de distribución de la frecuencia presentado en la columna final permite realizar comparaciones con distribuciones en otros grupos de mujeres. Los porcentajes no necesitan ser muy precisos. No hay necesidad de presentar la fracción 56 de 1.250 como 4,48% ya que 4,5% será suficiente en la mayoría de las situaciones. Al presentar los porcentajes de las dis-

Tabla 3.1.
Distribución de la altura en una muestra de 1.250 mujeres que acudieron a una clínica para cribaje de cáncer de mama: datos hipotéticos.

Altura (cm)	Número de mujeres	Porcentaje
145-149	75	6,0
150-154	153	12,2
155-159	261	20,9
160-164	323	25,8
165-169	201	16,1
170-174	144	11,5
175-179	91	7,3
180-184	2	0,2
Total	1.250	100,0

tribuciones de frecuencia debe indicarse siempre el número total de sujetos en los que se basa la distribución (1.250 en nuestro ejemplo). Por ejemplo, puede ser engañoso indicar que un 20% de las mujeres tenían una altura entre 155 y 159 cm si sólo fueron medidas 5 mujeres en total.

No existen normas universales sobre cómo agrupar los datos. Como guía aproximada, el número de grupos debe ser de 5 a 20, dependiendo del número de observaciones. Si el intervalo elegido para agrupar los datos es demasiado amplio se perderán demasiados detalles, mientras que si es demasiado estrecho la tabla será difícil de interpretar. Todos los intervalos deben tener la misma anchura aunque las categorías de los extremos pueden ser abiertas (p.e. ≥ 180 cm). No debe haber espacios entre los grupos. La tabla debe rotularse de modo que esté claro cómo se clasifican las observaciones que se encuentran en los extremos.

Una distribución de frecuencia puede representarse gráficamente mediante un histograma como se muestra en la Figura 3.1. En este tipo de gráfico el número (o porcentaje) de observaciones se representa gráficamente para diferentes valores, o grupos de valores, de la variable en estudio (altura, por ejemplo). En un histograma, el *área* del rectángulo es la que representa la frecuencia, no su altura – la escala vertical se mide en frecuencia por unidad de valor y la escala horizontal se mide en unidades de valor. Cuanto mayor sea el tamaño de la muestra, más refinado debe elegirse el intervalo de agrupación, de modo que el histograma se suaviza y se parece más a la distribución de la población total. En el límite, cuando la anchura de los intervalos es tan pequeña que prácticamente corresponde a una sola unidad, el diagrama resultante sería una curva lisa.

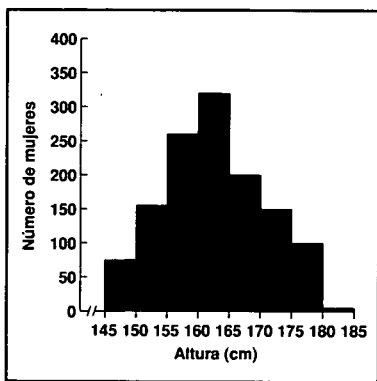


Figura 3.1.
Histograma que muestra la distribución de altura en 1.250 mujeres que acudieron a una clínica para cribaje de cáncer de mama (datos mostrados en la Tabla 3.1).

Resumiendo datos cuantitativos

Las tablas y los gráficos son importantes para examinar los datos. Pero también es deseable poder resumir los datos en forma numérica. Los dos aspectos más importantes de una distribución son normalmente el valor central y la dispersión sobre este valor.

Medidas del valor central

(1) La medida del valor central de una distribución utilizada con mayor frecuencia es la *media aritmética*. Ésta es la suma de las observaciones divi-

divida entre n , el número de observaciones. Por ejemplo, si los pesos (en kg) de 8 mujeres sanas son:

45,3 49,8 50,5 60,7 65,2 69,4 73,2 75,9

entonces la media aritmética de los pesos de estas mujeres es:

$$(45,3 + 49,8 + 50,5 + 60,7 + 65,2 + 69,4 + 73,2 + 75,9) / 8 = 490 / 8 = 61,25 \\ = 61,3 \text{ kg}$$

(2) La *mediana* es otra medida del valor central utilizada habitualmente. Es el valor que divide la distribución por la mitad cuando las observaciones se han ordenado. Así, la mediana es la observación central. Si existe un número par de observaciones, la mediana se calcula como la media de los 2 valores centrales. Como expresión general para calcular la mediana es:

Mediana = valor de la observación en la posición $(n+1) / 2^{\text{ava}}$, cuando las observaciones están ordenadas.

Donde n es el número total de observaciones.

En el ejemplo anterior, el valor de la mediana es el valor 4,5º, es decir, la media de los valores 4º y 5º, $(60,7 + 65,2) / 2 = 63,0$ kg.

La elección de que medida usar depende de la naturaleza de los datos y del objetivo del análisis. Con frecuencia la media es la medida preferida del valor central porque considera cada observación y es fácil de tratar en los tipos más comunes de análisis estadísticos. Su principal inconveniente es que puede ser afectada por valores extremos (las observaciones individuales que son extremas en comparación con la mayoría de las observaciones, y cuya inclusión o exclusión cambia el resultado considerablemente).

La mediana es una medida descriptiva útil cuando la presencia de valores muy extremos en los datos hace que la media no sea representativa de la mayoría de los datos. También es especialmente útil si no se registran observaciones cuando se encuentran por encima o por debajo de un determinado nivel; en estas circunstancias, no podemos calcular la media, pero podemos calcular la mediana si tenemos valores definidos para más de la mitad de los sujetos. Por ejemplo, sólo puede calcularse el tiempo medio de supervivencia de un grupo de pacientes con cáncer una vez han muerto todos. Pero es posible calcular la mediana de supervivencia aunque la mitad de los pacientes todavía estén vivos. El principal inconveniente de esta medida es que agrupa los datos pero no utiliza sus valores individuales.

Cuando la forma de una distribución es prácticamente simétrica alrededor del eje central, la media y la mediana son aproximadamente iguales, como vimos con los datos del ejemplo anterior (peso).

Medidas de variación

Además de una medida del valor central de una distribución, también es útil tener una idea de la variación o dispersión de los valores alrededor de este valor central. Existen varias medidas potenciales:

(1) El *rango* es el intervalo entre los valores mayor y menor. La principal ventaja de esta medida es que es muy simple de calcular. El principal inconveniente es que se basa en sólo dos observaciones extremas y no da una idea de cómo se disponen las observaciones entre estas dos.

(2) Los tres valores que dividen una distribución en cuartos se denominan *cuartiles*. El 25% de las observaciones se encuentra por debajo del cuartil inferior, un 50% por debajo del cuartil medio y un 75% por debajo del cuartil superior. El cuartil medio es la mediana. La distancia entre el cuartil inferior y el cuartil superior se denomina intervalo inter-cuartil y a veces se usa para describir la variabilidad. El *intervalo inter-cuartil* contiene un 50% de las observaciones.

Asimismo, los *percentiles* son valores que dividen la distribución en porcentajes. El percentil 50 corresponde a la mediana. Los percentiles 25 y 75 corresponden a los cuartiles inferior y superior, respectivamente.

Una manera semi-gráfica simple pero útil de resumir los datos utilizando percentiles es la representación en diagrama de cajas 'box-and-whisker'. La Figura 3.2 muestra un diagrama de cajas 'box-and-whisker' para los datos de las alturas de la Tabla 3.1. La caja indica los cuartiles inferior y superior y la línea central es la mediana. Los puntos en los extremos de las barras son los valores de los percentiles 2,5 y 97,5.

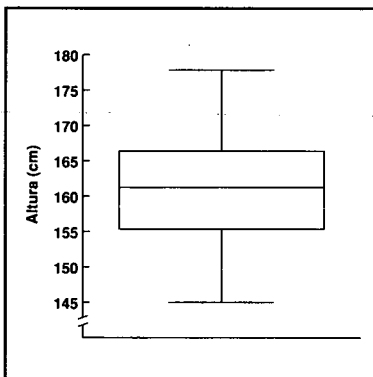


Figura 3.2. Diagrama de cajas 'box-and-whisker' para las alturas de la muestra de 1.250 mujeres presentadas en la Tabla 3.1.

(3) La *desviación estándar (DE)* es la medida de dispersión de los valores alrededor de la media utilizada con mayor frecuencia. Si los valores de una variable no varían mucho dentro de una población, las observaciones se encontrarán cerca y alrededor de la media, mientras que si existe una variación substancial, las observaciones estarán ampliamente dispersas alrededor de la media.

Esta variabilidad o *varianza* puede medirse en términos de cuán lejos se encuentran las observaciones de la media, es decir, cuán lejos, de promedio, se *desvía* cada observación de la media. Esto se ilustra en la Figura 3.3 para los ocho pesos de las ocho mujeres sanas del ejemplo anterior. Las desviaciones de la media se muestran por las líneas d_1, d_2, \dots, d_8 . En primer lugar, se calcula la suma de estas desviaciones; no obstante, la suma de las desviaciones de la media aritmética es, por definición, cero, es decir, las desviaciones negativas anulan las desviaciones positivas. No obstante, al calcular la dispersión de los valores alrededor de la media aritmética, es irrelevante si las desviaciones son positivas o negativas; sólo es interesante su magnitud numérica absoluta. Por tanto, para evitar obtener un cero al sumar las desviaciones, estas se elevan al cuadrado antes para eliminar los signos negativos. La media de estas desviaciones elevadas al cuadrado se denomina la *varianza*.

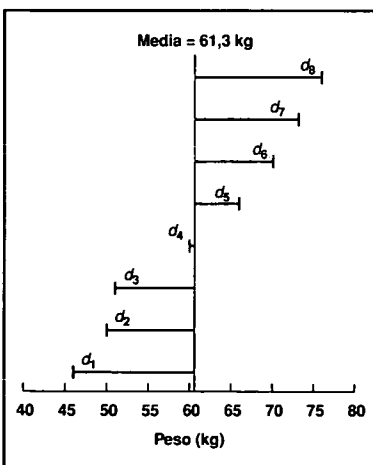


Figura 3.3. Desviaciones de la media de los pesos de 8 mujeres: datos hipotéticos.

$$\text{Varianza} = (d_1^2 + d_2^2 + \dots + d_8^2) / n$$

La *varianza* es una medida de la dispersión media muy útil, pero es difícil de interpretar porque es el cuadrado de las unidades utilizadas. En nues-

tro ejemplo, la varianza de peso se expresará en kg^2 . Normalmente es más conveniente expresar la variación en términos de las unidades originales no elevadas al cuadrado (en este caso kg). Así, necesitamos obtener la *raíz cuadrada de la varianza*. Esta se denomina *desviación estándar (DE)*. En este ejemplo, la DE es igual a 10,8 kg .

Una desviación estándar pequeña indica que la mayoría de los valores se encuentran próximos a la media, mientras que una DE grande indica que muchos valores se encuentran lejos de la media, es decir, cuanto más varían los valores en una población, mayor será la desviación estándar.

Como norma general, considerando una distribución prácticamente simétrica y en forma de campana, con una cúpula de valores centrales y dos colas en los extremos (es decir, lo que los estadísticos denominan una 'Distribución normal'), la media es el valor central y la desviación estándar es una medida de la dispersión, en la que una desviación estándar de los dos lados de la media incluye aproximadamente el 70% de la observación y dos desviaciones estándar incluye aproximadamente el 95% (Figura 3.4).

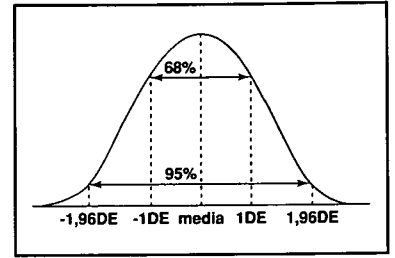


Figura 3.4. Distribución normal (DE: desviación estándar)

3.3.2 Datos categóricos

Los valores de las variables categóricas (también denominadas cualitativas) representan atributos más que determinaciones reales de alguna cantidad. Algunos ejemplos de variables categóricas son: sexo (hombre/mujer), estado civil (soltero/casado/divorciado/viudo), uso de anticonceptivos orales (alguna vez/nunca), país de nacimiento (Colombia, España, etc.).

Existen diferentes tipos de variables categóricas. Si la variable puede tomar sólo dos valores diferentes, la variable categórica se denomina binaria (p.e. sexo). A veces las diferentes categorías de una variable pueden ordenarse en algún tipo de escala (p.e. la gravedad del dolor se puede clasificar como leve, moderado y grave). En esas circunstancias, la variable se denomina una variable categórica ordinal.

Presentación de datos categóricos

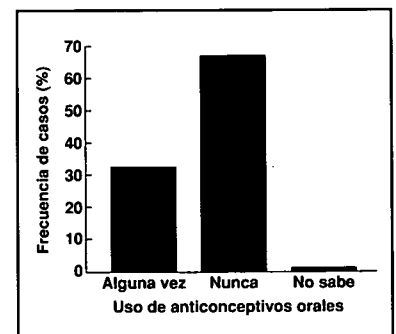
Al igual que con las variables cuantitativas, podemos presentar la distribución de frecuencia de una variable categórica como el número de veces que aparece cada valor (categoría) en el grupo de sujetos en estudio. Consideremos el Ejemplo 3.1.

Los diagramas de barras se utilizan habitualmente para representar la distribución de variables categóricas (Figura 3.5). En este tipo de gráfico, las frecuencias se representan por las longitudes de las barras.

Resumen de datos categóricos

Para los datos categóricos, los porcentajes son las únicas medidas básicas de resumen que pueden calcularse (ver Tabla 3.2). Son especialmente útiles para establecer comparaciones entre diferentes grupos.

Figura 3.5. Distribución de los casos de cáncer de cuello uterino según uso de anticonceptivos orales (datos de Bosch *et al.*, 1992).



Ejemplo 3.1. En España y Colombia se realizó un estudio para evaluar la relación entre el cáncer de cuello uterino y la exposición al virus del papiloma humano (VPH), comportamiento sexual y reproductivo, uso de anticonceptivos orales, cribajes practicados de cáncer de cuello uterino y hábito tabáquico. El estudio incluyó 436 casos de carcinoma escamoso confirmados histológicamente y 387 controles elegidos al azar entre la población general que generó los casos. Cada participante respondió a un cuestionario estructurado que recogía datos sobre un gran número de variables: edad, número de parejas sexuales, nivel cultural, hábito de fumar, etc. (Bosch et al., 1992). La Tabla 3.2. muestra la distribución de los casos de cáncer de cuello uterino según nivel de estudios y uso de anticonceptivos orales:

Tabla 3.2.

Distribución de los casos de cáncer de cuello uterino según nivel cultural y uso de anticonceptivos orales^a.

Características basales	Número de casos	Porcentaje
Educación (escolarización)		
Alguna vez	317	72,7
Nunca	119	27,3
Total	436	100,0
Uso de anticonceptivos orales		
Alguna vez	141	32,4
Nunca	291	66,7
No sabe	4	0,9
Total	436	100,0

^a Datos de Bosch et al. (1992).

3.3.3 Dos variables – cuantitativas o categóricas

Hasta ahora hemos considerado la distribución de la frecuencia de una variable en un único grupo de sujetos. Con frecuencia, queremos comparar las distribuciones de frecuencia entre dos o más grupos definidos por otra variable – por ejemplo, la distribución del uso de anticonceptivos orales entre los casos de Colombia y de España. Así podemos examinar la asociación entre dos variables categóricas, uso de anticonceptivos orales y país de residencia. Una manera de hacer esto es representando los datos como en la Tabla 3.3.

Tabla 3.3.

Distribución de los casos de cáncer de cuello uterino según el uso de anticonceptivos orales (filas) y el país de residencia (columnas)^a.

Uso de anticonceptivos orales	País de residencia (número (%))	
	Colombia	España
Alguna vez	77 (41,4)	64 (25,6)
Nunca	109 (58,6)	182 (72,8)
No sabe	0 (0)	4 (1,6)
Total	186 (100,0)	250 (100,0)

^a Datos de Bosch et al. (1992)

Al considerar dos variables simultáneamente es útil clasificarlas de acuerdo con el objetivo de la investigación, bien como variables explicativas o

como variables de respuesta. Las variables explicativas se refieren a las características de los sujetos (exposiciones) que deben utilizarse para explicar algo de la variable respuesta, que es el resultado de interés.

La Tabla 3.3 muestra la distribución de la variable respuesta (uso de anticonceptivos orales) según cada categoría de la variable explicativa (país de residencia). En este ejemplo, es adecuado calcular los porcentajes de columna con objeto de mostrar la distribución de la variable respuesta.

Estos datos también pueden presentarse como en la Tabla 3.4. En este caso, es adecuado calcular los porcentajes de las filas.

País de residencia	Uso de anticonceptivos orales (número (%))			Total
	Alguna vez	Nunca	No sabe	
Colombia	77 (41,4)	109 (58,6)	0 (0)	186 (100,0)
España	64 (25,6)	182 (72,8)	4 (1,6)	250 (100,0)

^a Datos de Bosch *et al.* (1992)

La distribución del uso de anticonceptivos orales entre los casos de cáncer de cuello uterino en los dos países puede representarse en un diagrama de dos barras unidas por los lados como en la Figura 3.6.

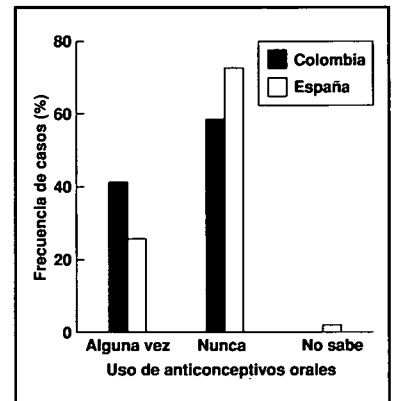
También podemos tabular las variables cuantitativas (edad, por ejemplo) agrupando los valores de la variable. La Tabla 3.5 muestra la distribución de los casos de cáncer de cuello uterino en el momento de su inclusión en el estudio en cada país.

La tabla muestra la distribución de la variable respuesta (edad) según cada categoría de la variable explicativa (país).

Las tabulaciones son útiles para mostrar la asociación entre variables categóricas. No son adecuadas para ilustrar la asociación entre variables cuantitativas (a no ser que se conviertan en variables categóricas como en la Tabla 3.5). Esto se comenta con más detalle en el Apartado 11.2.1; por ahora limitaremos la discusión a las representaciones gráficas de la asociación entre dos variables cuantitativas. Como ejemplo, se han representado en un

Tabla 3.4. Distribución de los casos de cáncer de cuello uterino por el uso de anticonceptivos orales (columnas), según el país de residencia (filas)^a.

Figura 3.6. Distribución de los casos de cáncer de cuello uterino según el uso de anticonceptivos orales y el país de residencia (datos de Bosch *et al.*, 1992).



Grupos de edad (años)	País de residencia (número (%))	
	Colombia	España
<30	15 (8,1)	7 (2,8)
30-39	48 (25,8)	41 (16,4)
40-44	27 (14,5)	30 (12,0)
45-54	50 (26,9)	61 (24,4)
≥55	46 (24,7)	111 (44,4)
Total	186 (100,0)	250 (100,0)

^a Datos de Bosch *et al.* (1992).

Tabla 3.5. Distribución de la edad de los casos de cáncer de cuello uterino según el país de residencia^a

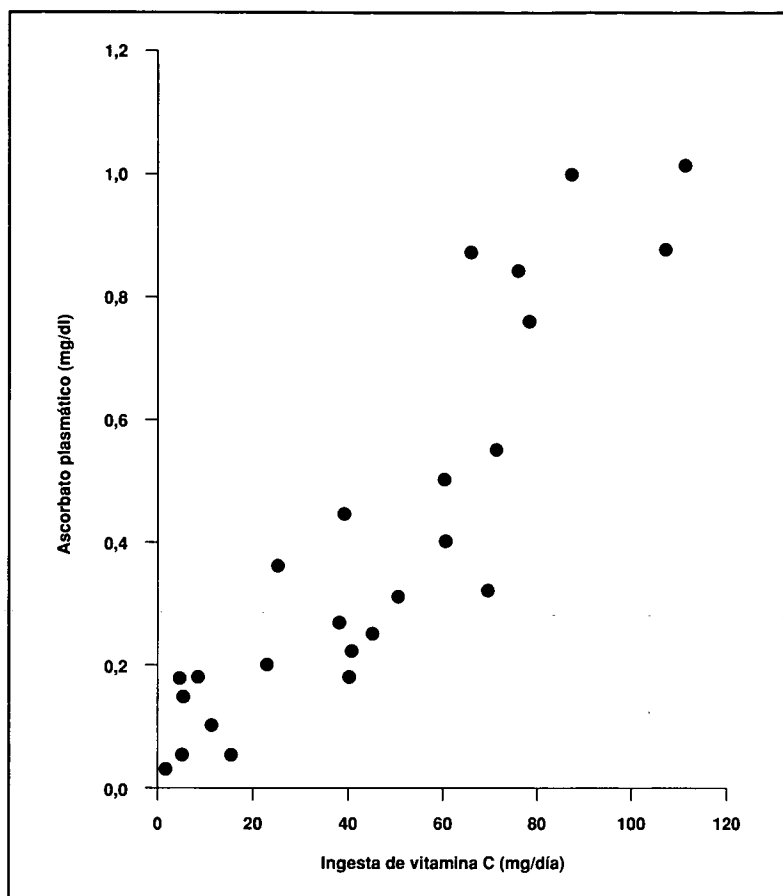


Figura 3.7. Niveles de ascorbato plasmático en 25 sujetos en estudio en relación con su ingesta de vitamina C: datos hipotéticos.

diagrama de dispersión (Figura 3.7) los datos de ingesta de vitamina C y los niveles en plasma en una muestra de 25 individuos. Los valores de cada variable se representan en un eje. Se utiliza un símbolo (con frecuencia un punto o una cruz) para representar cada individuo en el gráfico.

3.4 Observaciones finales

Los análisis de los datos deben iniciarse siempre mediante la elaboración de tablas, gráficos y estadísticas básicas que nos permitan explorar los datos. No obstante, la presentación de los mismos debe ser sensata y no prestarse a confusiones.

Normas generales para el diseño de tablas:

- (1) Todas las tablas deben tener un título que indique claramente su contenido. El lector debe poder determinar de forma precisa y sin dificultad lo que está tabulado.
- (2) Los títulos de las columnas y filas deben ser breves pero explicativos.
- (3) Se debe incluir la fuente de los datos de modo que puedan comprobarse las fuentes originales.
- (4) Si se incluyen unidades de medida, éstas deben indicarse claramente.
- (5) Se deben proporcionar los totales. Éstos son útiles para el lector para precisar cuántas observaciones hay en la tabla y como un medio de comparación y referencia con los datos presentados en otras tablas y en el texto.
- (6) Si se representan los porcentajes, la base de un porcentaje debe indicarse claramente. En una tabla compleja en la que la base no está indicada, el lector no está seguro de cuál o cómo los porcentajes suman 100 por ciento. Una tabla así puede malinterpretarse fácilmente. Una mirada a la posición del valor del 100 por ciento casi siempre informa al lector inmediatamente de cómo se obtuvieron los porcentajes de la tabla.
- (7) Las aproximaciones y omisiones pueden explicarse en notas a pie. Si se excluye información de la tabla, debe indicarse su número claramente.
- (8) Las tablas se utilizan para expresar los resultados de una manera más clara y concisa de lo que sería posible en el texto. Con frecuencia es necesario cierto grado de complejidad para ahorrar espacio. No obs-

tante, debe evitarse comprimir demasiada información en una única tabla. Dos o tres tablas sencillas pueden ser mejor que una tabla muy grande y compleja.

Normas generales para el diseño de gráficos:

- (1) Un gráfico debe tener una leyenda autoexplicativa.
- (2) Un gráfico debe ayudar al lector a entender los datos. No deben estar atestados con demasiados detalles.
- (3) Los ejes deben estar claramente titulados y las unidades de medida indicadas. Es importante para el lector poder entender con precisión lo que se está ilustrando y en qué unidades.
- (4) Las escalas utilizadas son extremadamente importantes. Las escalas deben empezar en cero; si no, debe indicarse claramente por un corte en el eje correspondiente.
- (5) Los gráficos en general son más fáciles de interpretar si la variable explicativa (exposición) se representa en el eje horizontal y la variable respuesta (resultado) en el eje vertical.
- (6) Evite gráficos con una impresión tridimensional debido a que pueden ser confusos (las tres dimensiones no se visualizan fácilmente).
- (7) Elija el tipo de gráfico que use la menor tinta posible sin pérdida de información. La Figura 3.8 presenta los mismos datos de dos formas

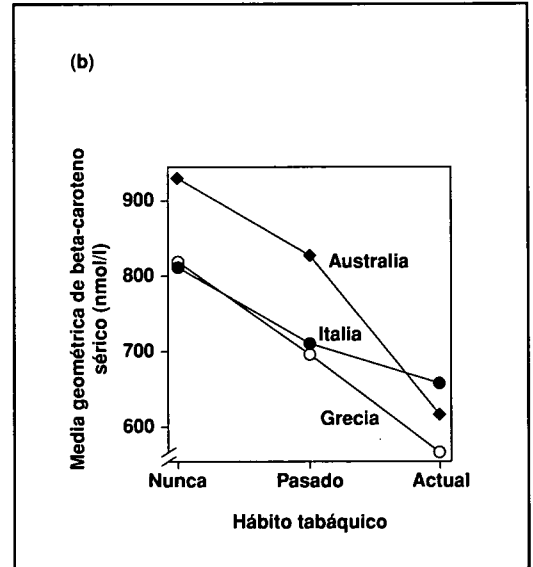
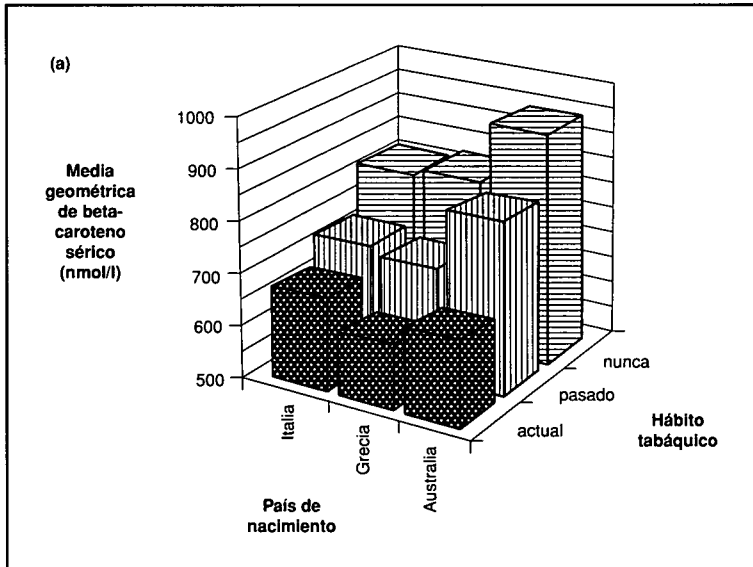


Figura 3.8.

Dos presentaciones diferentes de gráficos sobre 750 sujetos según su país de nacimiento, hábito tabáquico y concentraciones séricas de beta-caroteno: a) gráfico de 'negocios'; b) gráfico 'científico' (reproducido con la autorización de Jolley, 1993. © The Lancet Ltd., 1993).

diferentes. En el gráfico (a) los datos son difíciles de interpretar debido a las columnas tridimensionales, múltiples escalas de resultados, las líneas de reticulado y el sombreado. La mayor parte de la tinta se utilizó para presentar aspectos que pueden obviarse sin pérdida de información. En el gráfico (b) se presentan los mismos datos en una forma mucho más simple y clara.

Normas generales para presentar medidas básicas de resumen:

- (1) Indicar siempre el número de observaciones en el que se basa la medida. Para respuestas binarias (A, B), indicar el porcentaje de A o B, pero no los dos.
- (2) Si se utiliza la mediana para medir el valor central de una distribución cuantitativa, indicar también los cuartiles (o el intervalo) inferior y superior.
- (3) Si se utiliza la media para medir el valor central de una distribución cuantitativa, indicar también la desviación estándar.

Cuadro 3.1. Puntos esenciales

- Después de editar los datos el investigador debe examinarlos utilizando tabulaciones, gráficos y medidas básicas de resumen.
- La elección de los métodos correctos para presentar y resumir los datos depende del tipo de datos obtenidos: cuantitativos o categóricos.
- La distribución de frecuencias de una variable cuantitativa puede expresarse en tablas o gráficamente en histogramas y puede resumirse indicando una medida del valor central y una medida de la dispersión de la distribución (es decir, la media aritmética con desviación estándar o mediana con intervalo inter-cuartil).
- La distribución de frecuencias de una variable categórica puede presentarse en tablas o gráficamente en diagramas de barras. Los porcentajes de los individuos en cada categoría son las únicas medidas resumen que pueden calcularse para este tipo de variable.
- Las tablas y los diagramas de barras pueden utilizarse para examinar la relación entre dos variables categóricas y los diagramas de dispersión para estudiar la asociación entre dos variables cuantitativas. No obstante, es importante decidir cuál de las dos variables es la variable explicativa y cuál es la variable respuesta antes de trazar una tabla o gráfico.
- Las tablas, gráficos y medidas básicas de resumen deben utilizarse de manera sensata para asegurar que proporcionen una representación exacta de los datos.

Lecturas adicionales

* En Altman (1991) y Bland (1987) se encuentran más detalles sobre las formas de presentar y resumir datos.

* Un libro excelente sobre métodos gráficos es el de Tufte (1983).

* Jolley (1993) presenta una visión crítica del uso (y maluso) de los gráficos en epidemiología.

Capítulo 4

Medidas de la aparición de enfermedad y de otros acontecimientos relacionados con la salud

4.1 Introducción

La investigación epidemiológica se basa en la capacidad de cuantificar la aparición de enfermedad (o de cualquier otro acontecimiento de interés relacionado con la salud) en poblaciones. Esto requiere una definición clara de:

- (1) Qué se entiende por caso, es decir, un individuo en una población que tiene la enfermedad, alteración de la salud o padece un acontecimiento (p.e. muerte) de interés.
- (2) La población de la que proceden los casos.
- (3) El periodo de tiempo en el cual se obtuvieron los datos.

4.1.1 Definiendo un caso – el numerador

La definición de 'caso' conlleva ciertas dificultades. La definición epidemiológica de caso no coincide siempre con la definición clínica usual y, en la mayoría de los casos, los epidemiólogos están forzados a confiar en las pruebas de diagnóstico que son menos invasivas y más baratas que la utilizadas normalmente por los clínicos. Es importante estandarizar la definición de caso. Por ejemplo, ¿deben los 'casos de cáncer' incluir sólo los confirmados histológicamente? ¿Deben incluir lesiones *in situ*? Para cánceres que se desarrollan en órganos pareados (p.e. mama, testículos, riñón), ¿se deben contar los casos como el número de personas o el número de órganos afectados? Los epidemiólogos que se dedican al estudio del cáncer también pueden estar interesados en medir la frecuencia de otros acontecimientos relacionados con la salud, así por ejemplo, un 'caso' podría ser alguien que utiliza anticonceptivos orales, fuma o utiliza una parte determinada de los servicios sanitarios.

Otra consideración importante para situaciones recurrentes no fatales (p.e. el resfriado común) es decidir si, para un individuo determinado, cada episodio repetido debe contarse como un caso o sólo el primero. En este capítulo, hemos supuesto que los individuos sólo pueden padecer un episodio de la situación de interés, pero las medidas de aparición presentadas aquí pueden modificarse para aceptar más de un episodio por individuo.

Se pueden identificar casos utilizando registros de enfermedad, sistemas de notificación, certificados de defunción, resúmenes de registros clínicos, revisiones de la población general, etc. Es importante asegurar que el nu-

merador incluya todos los casos que aparecen en la población en estudio y que no se incluya ningún caso de otra población. Por ejemplo, si estamos interesados en determinar la aparición de una enfermedad en una localidad en particular, todos los casos que aparecieron entre sus residentes deben incluirse en el numerador aunque se hubieran diagnosticado en otro lugar. Por el contrario, deben excluirse los casos diagnosticados en esta localidad que normalmente viven en otro lugar.

4.1.2 Definiendo la población a riesgo – el denominador

El conocimiento del número de casos que aparecieron en una población en particular tiene de por sí poca utilidad para los epidemiólogos. Por ejemplo, saber que se produjeron 100 casos de cáncer de pulmón en la ciudad A en comparación con 50 casos en la ciudad B, no nos permite concluir que el cáncer de pulmón es más frecuente en la ciudad A que en la B. Para comparar la frecuencia de cáncer de pulmón en estas dos poblaciones, también es necesario conocer el tamaño de las poblaciones de las que proceden los casos (es decir, el denominador).

Es importante definir claramente la población a riesgo. Ésta puede ser la de los residentes de una ciudad en particular, la población de todo un país o la población de captura de un hospital. La definición debe excluir todos los que no son residentes habituales en esta área. También debe excluir a todos aquellos que no presentan riesgo de desarrollar el acontecimiento en estudio, aunque algunas veces esto puede no ser posible. Por ejemplo, las mujeres sometidas a histerectomía deben excluirse del denominador si se quiere cuantificar la aparición de cáncer de cuello uterino en una población determinada, ya que estas mujeres no presentan el riesgo de desarrollar este cáncer. No obstante, raramente están disponibles los datos necesarios para excluir a estas mujeres y normalmente se incluyen todas las mujeres en el denominador.

4.1.3 Período de tiempo

Es difícil interpretar medidas de la aparición de eventos relacionados con la salud a no ser que se haya establecido claramente el periodo de tiempo durante el cual la población estaba a riesgo (y se reunieron los casos). Esto es debido a que la mayoría de las enfermedades no se producen de manera constante en el tiempo. Por ejemplo, la aparición de cáncer de pulmón en la mayoría de los países occidentales fue mucho menor en los primeros años de este siglo que ahora.

4.2 Medidas de la aparición

Existen dos medidas principales de aparición de una enfermedad: prevalencia e incidencia.

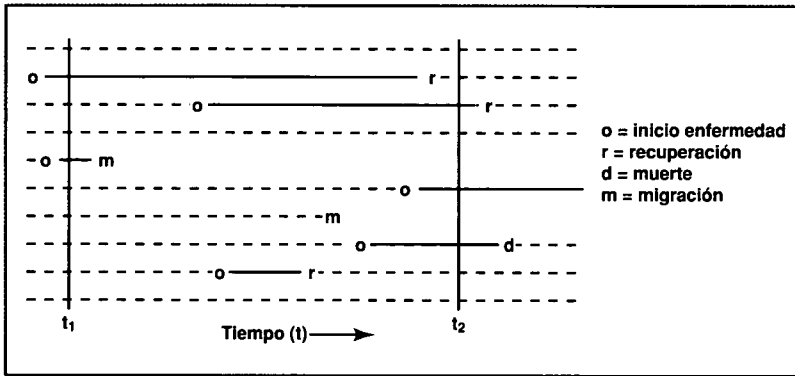
4.2.1 Prevalencia

La prevalencia puntual es la proporción de casos existentes (anteriores y nuevos) en una población en un único punto en el tiempo.

$$\text{Prevalencia puntual} = \frac{\text{N}^\circ \text{ de casos existentes en una población definida en un punto en el tiempo}}{\text{N}^\circ \text{ de personas en la población definida en el mismo punto en el tiempo}}$$

Esta medida se denomina prevalencia puntual^a porque se refiere a un punto único en el tiempo. Con frecuencia simplemente se describe como prevalencia.

Ejemplo 4.1. Cada línea en el diagrama mostrado en la Figura 4.1. representa un sujeto en una población particular. Algunos sujetos desarrollaron la enfermedad de interés y se recuperaron o murieron de ella. Otros abandonaron la población y se fueron a vivir a otro sitio. Debido a estos cambios dinámicos, la magnitud de la prevalencia varía de un punto en el tiempo a otro.



Prevalencia en el tiempo $t_1 = \frac{2}{10} = 0,20 = 20\%$
Prevalencia en el tiempo $t_2 = \frac{3}{8} = 0,38 = 38\%$

Al igual que cualquier proporción, la prevalencia no tiene unidades de tiempo. No obstante, el punto en el tiempo al que se refiere debe especificarse siempre. El término 'tasa de prevalencia' se utiliza con frecuencia en lugar de 'prevalencia'. Esto no es correcto ya que la prevalencia, por definición, es una proporción y no una tasa (ver Apartado 4.2.2).

Ejemplo 4.2. En 1985 se llevó a cabo un estudio en una pequeña localidad para determinar la prevalencia del uso de anticonceptivos orales entre mujeres de 15-44 años. Todas las mujeres entre estas edades residentes en la localidad se entrevistaron sobre el uso actual de anticonceptivos orales. La prevalencia del uso de anticonceptivos orales en esta localidad en 1985 entre mujeres de 15-44 años fue 0,5 (= 50%).

Puede ser difícil definir un caso de cáncer prevalente. Los registros de cáncer generalmente suponen que una vez un individuo ha sido diagnosti-

^a La prevalencia de período es una variación que representa el número de personas consideradas como casos en cualquier momento durante el período (corto) especificado dividido por el número total de personas en esta población durante ese tiempo. Esta medida se utiliza raramente en epidemiología del cáncer. Se utiliza cuando la enfermedad es recurrente y no fatal. Un ejemplo de prevalencia periódica sería la proporción de mujeres que han utilizado anticonceptivos orales en algún momento durante el período de 12 meses anterior al día de la revisión.

Figura 4.1. Cambios en el tiempo en el estado de salud y en la migración de algunos miembros de una población particular y cómo estos cambios afectan a la prevalencia de la enfermedad en esta población.

cado de cáncer, éste sigue siendo un caso prevalente hasta la muerte (ver Apartado 17.6.1). No obstante, esta suposición no siempre es correcta, ya que algunas personas a las que se les diagnostica cáncer pueden sobrevivir durante mucho tiempo y morir por otra causa sin evidencias de que el cáncer haya recurrido.

La prevalencia es la única medida de la aparición de la enfermedad que puede obtenerse de estudios transversales (Capítulo 10). Mide la carga de la enfermedad en una población. Esta información es útil para los planificadores de salud pública y administradores que pretendan anticipar los recursos sanitarios y los servicios requeridos por una población, adaptándose a las necesidades reales.

4.2.2 Incidencia

El número de casos de «enfermedad» presentes en una población en un punto en el tiempo depende de la frecuencia de aparición e identificación de los casos nuevos y de la duración media de la enfermedad (es decir, tiempo hasta la recuperación o la muerte). Como consecuencia, la prevalencia puede variar de una población a otra solamente debido a variaciones en la duración de la enfermedad. Por tanto no es una medida muy útil si se está interesado en establecer y cuantificar los determinantes de la enfermedad. En esta última situación, se está más interesado en medir el flujo de casos nuevos generado a partir de individuos sin la enfermedad.

A diferencia de la prevalencia, la incidencia cuantifica el número de casos nuevos de enfermedad que se desarrollan en una población de individuos a riesgo durante un intervalo de tiempo especificado. Pueden calcularse tres medidas distintas de incidencia: riesgo, *odds* de la enfermedad y tasa de incidencia.

Riesgo

El riesgo es la proporción de personas en una población inicialmente libres de enfermedad que desarrollan la enfermedad dentro de un intervalo de tiempo especificado:

$$\text{Riesgo} = \frac{\text{Nº de casos nuevos de enfermedad procedentes de una población definida en un período de tiempo determinado}}{\text{Nº de personas sin la enfermedad al inicio de este período de tiempo en esta población}}$$

Tanto el numerador como el denominador incluyen sólo aquellos individuos que, al inicio del periodo, no tienen la enfermedad y por tanto están a riesgo de desarrollarla. Esta medida de la incidencia puede interpretarse como la probabilidad media, o riesgo, de que un individuo desarrolle una enfermedad durante un periodo de tiempo especificado.

En la bibliografía epidemiológica se utilizan a menudo otros términos para indicar riesgo, tales como riesgo de incidencia y tasa de incidencia.

Al igual que cualquier proporción, el riesgo no tiene unidades. No obstante, su precisión aumenta con la duración del seguimiento y por tanto el periodo de tiempo al que se relaciona debe especificarse siempre claramente, como en el Ejemplo 4.3.

Ejemplo 4.3. *A principios de 1981 se reclutó un grupo de 5.000 mujeres sanas de 45-75 años a las que se siguió durante 5 años. Durante este periodo se detectaron 20 casos nuevos de cáncer de mama. Por tanto, el riesgo de desarrollar cáncer de mama en esta población durante este periodo de 5 años fue de $20 / 5.000 = 0,4\%$.*

El riesgo es una medida de aparición de la enfermedad utilizada habitualmente para cuantificar la supervivencia de un grupo de sujetos como en el Ejemplo 4.4.

Ejemplo 4.4. *En una población determinada en 1971 se diagnosticaron un total de 13.264 casos de cáncer de pulmón en hombres. Estos casos se siguieron durante 5 años. Al final de este periodo de seguimiento, sólo 472 casos estaban vivos todavía. La probabilidad de sobrevivir durante este periodo de 5 años fue de $472 / 13.264 = 3,6\%$. Así, la probabilidad de morir durante este periodo fue del $100\% - 3,6\% = 96,4\%$. Estas medidas son riesgos porque representan la proporción de casos de cáncer de pulmón que todavía estaban vivos (o murieron) al final del periodo de seguimiento entre todos los casos diagnosticados al inicio del estudio. En estos cálculos hemos supuesto que todos los individuos se siguieron durante todo el periodo de 5 años (o hasta su muerte si era anterior).*

Las medidas en el Ejemplo 4.4 con frecuencia se denominan 'tasas' de supervivencia o fatalidad, lo que es incorrecto, ya que, por definición, son proporciones y no tasas (ver más adelante en este apartado). Estas dos medidas se seguirán comentando en el Capítulo 12.

Odds de la enfermedad

Una segunda medida de la incidencia es la *odds* de la enfermedad que es igual al número total de casos dividido por el número total de personas que seguían sin presentar la enfermedad durante el periodo de estudio.

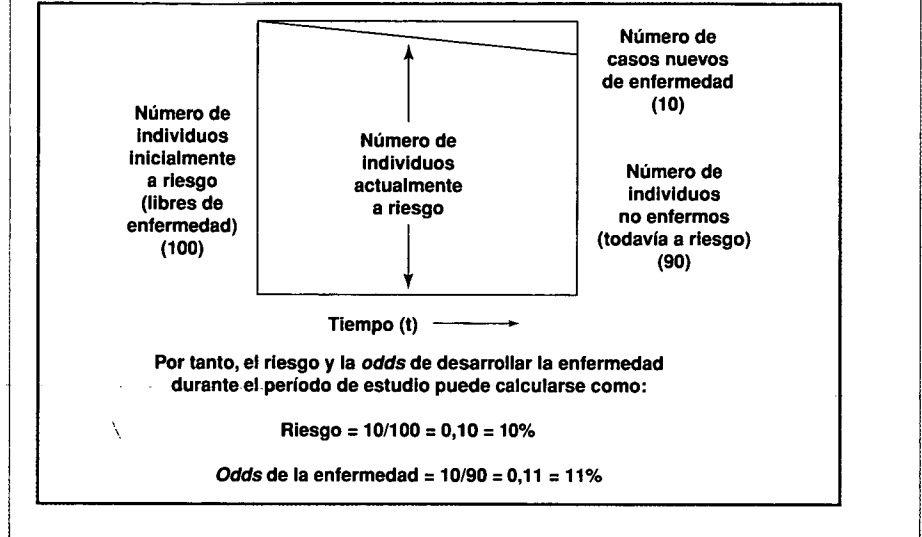
$$\text{Odds} = \frac{\text{N}^\circ \text{ de casos nuevos de enfermedad procedentes de una población definida en un período de tiempo determinado}}{\text{N}^\circ \text{ de personas que seguían sin la enfermedad durante este período de tiempo en esta población}}$$

Esta medida es una razón entre la probabilidad de adquirir la enfermedad y la probabilidad de no adquirirla durante un periodo de tiempo determinado. Por tanto, también puede expresarse como:

$$\text{Odds de la enfermedad} = \text{riesgo} / (1 - \text{riesgo})$$

Ejemplo 4.5. Supongamos que 100 individuos no enfermos se siguen durante un determinado periodo de tiempo. Al final de este periodo 10 habían desarrollado la enfermedad en estudio (Figura 4.2.)

Figura 4.2. Seguimiento de los 100 individuos no enfermos mencionados en el Ejemplo 4.5.



El riesgo y la *odds* de la enfermedad utilizan el mismo numerador (número de casos nuevos) pero denominadores diferentes. En el cálculo del riesgo, el denominador es el número total de individuos sin la enfermedad al inicio del periodo de estudio mientras que en el cálculo de la *odds* de la enfermedad es el número de individuos que siguieron sin la enfermedad al final del periodo de estudio.

Tasa de incidencia

Los cálculos de riesgo y *odds* de la enfermedad suponen que toda la población a riesgo al inicio del estudio se ha seguido durante el periodo de tiempo especificado. No obstante, a menudo algunos participantes se incluirán en el estudio unos años después de su inicio y algunos se perderán durante el seguimiento antes de su fin (población dinámica). En estos casos, la longitud del seguimiento no será uniforme para todos los participantes. Además, para los individuos que desarrollan la enfermedad, ninguna de las dos medidas de la incidencia consideradas hasta ahora tiene en cuenta el momento en que apareció la enfermedad.

Para considerar estas duraciones variables del seguimiento, el denominador puede calcularse de modo que represente la suma de los tiempos contribuidos por cada individuo a riesgo, es decir, la suma del tiempo que cada persona permanecía bajo observación y a riesgo de convertirse en un caso. Esto se conoce como persona-tiempo a riesgo, expresando el tiempo en unidades de medida apropiadas, p.e. personas-años (con frecuencia abreviado p-as).

El Ejemplo 4.6 ilustra el cálculo de persona-tiempo a riesgo utilizando un grupo hipotético de nueve personas.

Ejemplo 4.6. Consideremos un grupo hipotético de nueve sujetos que se siguieron desde el principio de 1980 al final de 1984. Los sujetos se incluyeron en el estudio a diferentes momentos como se muestra en la Figura 4.3. Tres sujetos – (2), (6) y (7) – desarrollaron la enfermedad durante el periodo de estudio y uno – (4) – se contactó por última vez a finales de 1983.

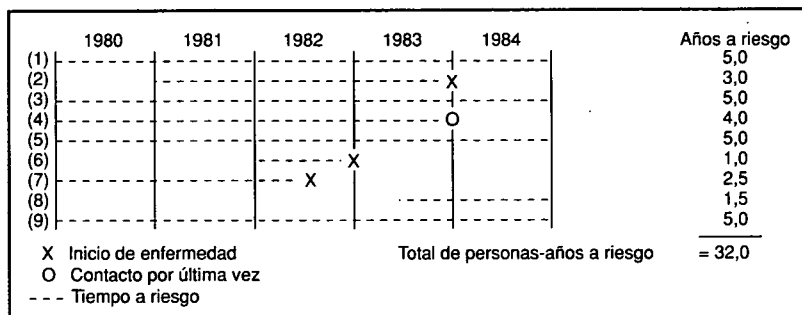


Figura 4.3.

Cálculo del tiempo a riesgo de cada individuo y el total de persona-tiempo a riesgo para los nueve sujetos del Ejemplo 4.6.

En el Ejemplo 4.6, el sujeto (1) se incluyó en el estudio a principios de 1980 y se siguió durante todo el periodo de estudio. Por tanto el sujeto estaba a riesgo de adquirir la enfermedad durante todo el periodo del estudio. El sujeto (4) también se incluyó en el estudio a su inicio pero se contactó con él por última vez a finales de 1983; por tanto, este sujeto sólo estuvo a riesgo durante 4 años. El sujeto (6) se incluyó en el estudio a principios de 1982 pero desarrolló la enfermedad a finales de este año; desde entonces, ya no estaba a riesgo (aquí suponemos que no hay recuperación posible de la enfermedad de interés). El total de personas-años a riesgo se calcula tomando solamente el tiempo de los individuos a riesgo.

El cálculo de la tasa de incidencia considera las posibles diferencias que pueda tomar el cálculo de la persona-tiempo a riesgo para cada individuo y es igual a:

$$\text{Tasa de incidencia} = \frac{\text{Nº de casos nuevos en una población definida en un período de tiempo determinado}}{\text{Persona-tiempo a riesgo total durante este período}}$$

Esta medida de la frecuencia de la enfermedad también se denomina densidad de incidencia o fuerza de morbilidad (o mortalidad). Al igual que el riesgo y la *odds*, el numerador de la tasa de incidencia es el número de casos nuevos en la población. El denominador, no obstante, se conoce como la suma del tiempo a riesgo de cada individuo. En el ejemplo hipotético anterior, la tasa de incidencia será igual a:

$$3/32 = 0,094 \text{ por personas-años o } 9,4 \text{ por } 100 \text{ personas-años}$$

Al presentar una tasa de incidencia, es esencial especificar las unidades de tiempo; esto es, si la tasa representa el número de casos por persona-día, persona-mes, persona-año, etc.

Las definiciones anteriores de riesgo, *odds* y tasa están ampliamente aceptadas, pero los términos riesgo y tasa se intercambian en gran parte de la bibliografía especialmente la menos reciente.

4.2.3 Relación entre prevalencia, tasa y riesgo

En el Apartado 4.2.2 se indicó que la prevalencia dependía de la incidencia y de la duración de la enfermedad. Cuando la incidencia y la duración de la enfermedad son estables y la prevalencia de la enfermedad es baja (p.e. cáncer), esta asociación puede expresarse como sigue:

$$\text{prevalencia} = \text{tasa de incidencia} \times \text{duración media de la enfermedad}$$

Ejemplo 4.7. Supongamos que cada año, en una población de 100.000 personas, se diagnostican un total de 50 casos nuevos de un tipo particular de cáncer. La duración media de (es decir, la supervivencia de) este cáncer se sabe que es de 4 años. Por tanto, la prevalencia de este tipo de cáncer en esta población es:

$$\text{prevalencia} = 0,0005 \text{ por persona-año} \times 4 \text{ años} = 0,2\%$$

El riesgo depende de la tasa de incidencia y de la duración del periodo a riesgo. El riesgo también se afecta por la mortalidad debida a enfermedades diferentes de la enfermedad de interés. Esto es debido a que algunos de los que murieron por otras enfermedades habrían desarrollado la enfermedad en estudio si hubieran sobrevivido. Si no se considera esta mortalidad por otras enfermedades y si la tasa de incidencia es constante durante el periodo a riesgo, puede aplicarse la siguiente relación:

$$\text{Riesgo} = 1 - \exp(- \text{tasa de incidencia} \times \text{duración del periodo a riesgo})$$

Donde 'exp' indica que la constante matemática $e = 2,72$ debe elevarse a la potencia de la expresión en paréntesis. Para enfermedades con una tasa de incidencia baja o cuando la duración del periodo a riesgo es corta, puede utilizarse la siguiente aproximación:

$$\text{Riesgo} = \text{tasa de incidencia} \times \text{duración del periodo a riesgo}$$

Esto se ilustra claramente en el Ejemplo 4.8.

Ejemplo 4.8. Supongamos que la tasa de incidencia de una enfermedad en particular en una determinada población es de 50 por 100.000 personas-año. El riesgo de que un individuo de esta población desarrolle esta enfermedad durante un periodo de 5 años (suponiendo que no intervengan otras causas de muerte) es:

$$\text{Riesgo durante 5 años} = 1 - \exp(-0,0005 \text{ por persona-año} \times 5 \text{ años}) = 0,0025 = 0,25\%$$

Se habría obtenido el mismo valor con las fórmulas simplificadas:

$$\text{Riesgo durante 5 años} = 0,0005 \text{ por persona-año} \times 5 \text{ años} = 0,0025 = 0,25\%$$

Consideremos ahora una enfermedad común con una tasa de incidencia de 300 por 1.000 personas-años:

$$\text{Riesgo durante 5 años} = 1 - \exp(-0,3 \text{ por persona-año} \times 5 \text{ años}) = 0,78 = 78\%$$

En este caso, la fórmula simplificada proporcionaría un resultado sin sentido:

$$\text{Riesgo durante 5 años} = 0,3 \text{ por persona-año} \times 5 \text{ años} = 1,5 = 150\%$$

(El riesgo es una proporción y por tanto nunca puede tener un valor superior a 1, o 100%).

4.3 Utilidad de las estadísticas vitales para medir la aparición de la enfermedad

Es posible estimar tasas a partir de datos obtenidos rutinariamente (p.e. datos de estadísticas vitales, datos de los registros de cáncer) aunque no se disponga de medidas directas de persona-tiempo a riesgo (Ejemplo 4.9). Una estimación del valor de persona-tiempo a riesgo de una población en un periodo dado se puede calcular como sigue:

Población existente en el punto medio del periodo de interés \times duración del periodo (en unidades de tiempo adecuadas, normalmente años)

Este método proporciona unas estimaciones adecuadas de persona-tiempo a riesgo siempre que la población permanezca estable durante este periodo.

Este método de estimar persona-tiempo a riesgo es adecuado para enfermedades raras tales como cáncer. No obstante, para enfermedades comunes será necesario utilizar aproximaciones más sofisticadas que excluyan del denominador a aquellos que ya tienen la enfermedad y por tanto ya no se consideran a riesgo.

Ejemplo 4.9. Supongamos que queremos estimar la incidencia de cáncer de estómago en hombres que viven en Cali (Colombia). El volumen VI de la serie *Incidencia del Cáncer en Cinco Continentes* [Cancer Incidence in Five Continents] (Parkin et al., 1992) proporciona datos sobre el número total de casos de cáncer de estómago producidos en Cali durante los años 1982-86 y sobre la población masculina total en 1984. El cálculo de la tasa de incidencia de cáncer de estómago a partir de estos datos se presenta a continuación:

Casos de cáncer de estómago en hombres, Cali, 1982-86 = 655

Población masculina, Cali, 1984 = 622.922

Personas-años a riesgo, 1982-86 = 5 (años) × 622.922 = 3.114.610 p-as

Tasa de incidencia anual media, Cali, 1982-86 = 655/3.114.610 = 21,03 por 100.000 p-as

Así, la tasa de incidencia media anual de cáncer de estómago en hombres que vivían en Cali durante los años 1982-86 fue de 21 por 100.000 personas-años.

La mayoría de los países desarrollados y muchos países en desarrollo tienen censo, normalmente uno cada diez años. El censo constituye el recuento basal de la población total. Como fuente de datos del denominador, los censos resultan incompletos para algunos subgrupos de la población (p.e. personas sin hogar, nómadas) o resultar obsoletos rápidamente si los cambios demográficos son importantes. Para resolver este problema, la mayoría de las oficinas censales proporcionan estimaciones del tamaño de la población entre los años de censo (años intercensales) basadas en las tasas de nacimiento, muerte y migración de la población. Si se dispone de estas estimaciones anuales de la población, éstas pueden considerarse como las mejores estimaciones disponibles de persona-tiempo a riesgo en cada año. Así, en el ejemplo anterior podría haberse utilizado la suma de las estimaciones de la población anual para los años 1982-86 (si están disponibles) para proporcionar una estimación de persona-tiempo a riesgo para todo el periodo de estudio.

4.3.1 Medidas crudas y específicas por estrato

Las medidas de la aparición de la enfermedad comentadas en el Apartado 4.2 pueden calcularse para una población entera o por separado para subgrupos específicos (estratos) de la población. En el primer caso se denominan medidas 'crudas' y en el segundo medidas 'específicas por estrato'. Por ejemplo:

$$\text{Tasa de incidencia cruda} = \frac{\text{N}^\circ \text{ de casos nuevos aparecidos en una población definida en un período de tiempo específico}}{\text{Total personas-años a riesgo en esta población durante este período de tiempo}} \times 100.000 \text{ por } 100.000 \text{ p-as}$$

Las tasas crudas se utilizan a menudo, en parte porque son medidas resumen, fáciles de interpretar y requieren poca información para ser calculadas. Las tasas crudas pueden ocultar el hecho de que subgrupos de la población muestren diferencias significativas en la incidencia. Por ejemplo, el riesgo de morir es diferente en los distintos grupos de edad. Esto debe considerarse al comparar tasas crudas de poblaciones diferentes, ya que diferencias en estas tasas pueden ser el reflejo de diferencias en la estructura de la población más que diferencias reales en la incidencia de la enfermedad para un mismo grupo de edad (véase Apartado 4.3.3).

Para entender algunos de los aspectos epidemiológicos de una enfermedad, es necesario el cálculo de tasas más detalladas, específicas para sexo y otros componentes demográficos tales como la edad. Por ejemplo, las tasas específicas por edad pueden calcularse como sigue:

$$\text{Tasa de incidencia específica de edad por 100.000 p-as} = \frac{\text{N}^{\circ} \text{ de casos nuevos aparecidos en un grupo de edad determinado, en una población definida y en un período de tiempo específico}}{\text{Personas-años a riesgo en este grupo de edad en la misma población y durante el mismo período de tiempo}} \times 100.000$$

El cálculo de persona-tiempo a riesgo se efectúa por separado para cada grupo de edad. La representación gráfica de las tasas específicas de la edad frente a la edad proporciona una curva de incidencia por edad que puede aportar claves importantes para la comprensión de la epidemiología de una enfermedad (véase Figura 4.5a). Cabe destacar que las tasas de cáncer deben ser siempre específicas de sexo, es decir, deben calcularse por separado para hombres y mujeres debido a que la incidencia de la mayoría de los tipos de cáncer difiere considerablemente entre sexos.

4.3.2 Cambios en la incidencia de la enfermedad con el tiempo

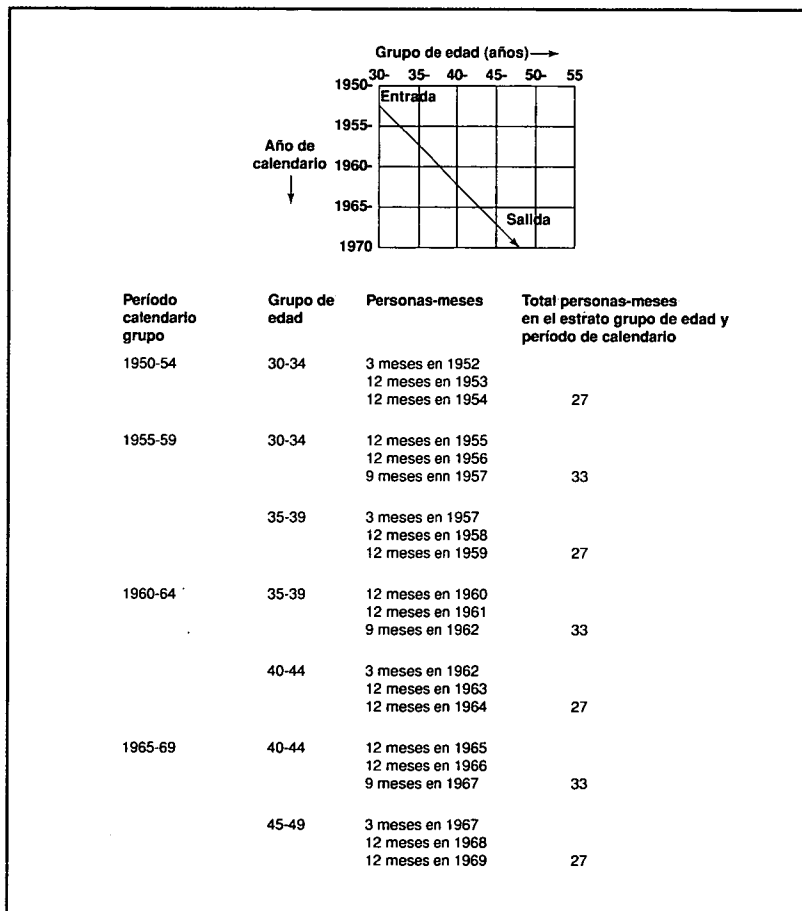
El riesgo de adquirir la enfermedad no sólo cambia con la edad sino también con el tiempo de calendario. Ambos aspectos deben tenerse en cuenta cuando se realiza un seguimiento de las personas a estudio. Esto se ilustra claramente en el Ejemplo 4.10.

La experiencia de toda la cohorte puede representarse en un diagrama formado por celdas de edad y tiempo de calendario o estratos (Figura 4.4). Este diagrama se denomina diagrama de Lexis y permite el examen simultáneo del seguimiento individual en relación con dos escalas de tiempo diferentes – edad y periodo de calendario. Una vez una persona se incluye en la cohorte, se desplaza en diagonal por el diagrama a medida que envejece y contribuye persona-tiempo a riesgo en varios estratos a medida que se desplaza por ellos. Las tasas específicas por estrato pueden calcularse dividiendo el número total de casos en cada estrato de edad-periodo calendario por el correspondiente valor de persona-tiempo a riesgo.

Ejemplo 4.10. Imaginemos un grupo de personas (o 'cohorte') con edades que oscilan entre 30 y 54 años que se siguieron desde 1950 hasta finales de 1969. Los sujetos en estudio contribuyeron persona-tiempo a riesgo desde el momento que se incluyeron en la cohorte hasta el final del estudio en 1969 (o hasta que cumplían 55 años, si se producía antes). La experiencia de uno de los sujetos a estudio se muestra en la Figura 4.4. Este sujeto cumplió 30 años el día que se incluyó en la cohorte el 1 de octubre de 1952. Tenía 47 años y 3 meses cuando el estudio terminó el 31 de diciembre de 1969.

Figura 4.4.

Diagrama de Lexis que muestra el seguimiento de un sujeto en estudio descrito en el Ejemplo 4.10 y el cálculo de su contribución personas-meses en cada estrato de edad y periodo de tiempo de calendario.



El diagrama de Lexis también puede utilizarse con datos de rutina para describir la incidencia de generaciones sucesivas incluso cuando no se dispone de los datos sobre la fecha de nacimiento real. La Tabla 4.1 presenta las tasas de mortalidad por cáncer de pulmón en hombres, en Inglaterra y Gales, 1941-78. Las columnas en esta tabla muestran cambios con la edad en las tasas de incidencia, y las filas, cambios de las tasas específicas por edad en el tiempo de calendario. En la tabla, las tasas pueden leerse en dos direcciones (por grupos de edad y por calendario), y las líneas diagonales representan cohortes de nacimiento sucesivas. La experiencia para las pri-

meras cohortes y las de nacimiento más reciente (las de los extremos en la tabla) contribuirán con muy pocos puntos de información.

Año de fallecimiento	Grupo de edad (años)																		
	0-	5-	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70-	75-	80- 85 y mayor		
1941-45	1	1	2	4	9	17	39	87	203	404	626	924	1073	1031	799	685	396		
1946-50	1	2	1	4	12	18	41	97	242	555	972	1375	1749	1798	1442	1130	791	497	
1951-55	1	0	1	2	8	14	37	100	250	584	1232	2018	2575	2945	2651	2087	1444	927	
1956-60	1	0	0	2	5	13	36	94	253	594	1254	2326	3330	3941	3896	3345	2271	1438	
1961-65	1	0	1	2	5	11	33	91	225	566	1226	2290	3673	4861	4994	4530	3423	2062	
1966-70	0	0	0	3	4	11	25	76	218	532	1165	2208	3703	5281	6223	5931	4578	3490	
1971-75	0	0	0	1	4	10	25	58	178	505	1074	2082	3552	5185	6834	7284	6097	4384	
1976-78	0	1	0	1	3	8	18	60	146	422	1056	1907	3384	5061	6782	7982	7395	5285	
							1951	1946	1941	1936	1931	1926	1921	1916	1911	1906	1901	1896	1891

Cohorte de nacimiento (diagonal)
 a Datos de OPCS, (1981).

Las diagonales de la Tabla 4.1 (desde la parte superior izquierda a la inferior derecha), por ejemplo, definen la experiencia de mortalidad por cáncer de pulmón para generaciones sucesivas en hombres que nacieron en el mismo periodo y por tanto envejecen juntos (p.e. un hombre de 40-44 años en 1941-45 tenía 45-49 en 1946-50, 50-54 en 1951-55, etc.). Para tener 40-44 años en 1941-45, este hombre podía haber nacido en cualquier momento entre Enero de 1896 (44 en 1941) y Diciembre de 1905 (40 en 1945). Las denominadas cohortes de nacimiento se identifican normalmente por su año central de nacimiento. Por ejemplo, la «cohorte de nacimiento en 1901» (o de manera más precisa la «cohorte de nacimiento en 1900/1») contiene a aquellos hombres nacidos durante el periodo de 10 años desde 1896 a 1905. La diagonal justo por encima de ésta, contiene las tasas correspondientes a la «cohorte 1896» (es decir, los hombres nacidos entre 1891 y 1900). Debido a que los años de nacimiento para cada cohorte se estiman a partir de la edad y del periodo calendario, las cohortes adyacentes se superpondrán inevitablemente, es decir, tienen años de nacimiento comunes. Cuando se dispone de los datos sobre el año de nacimiento exacto no es necesario hacer estas estimaciones y por tanto no existe superposición entre cohortes de nacimiento sucesivas.

Por tanto, los análisis de cohortes de nacimiento utilizan las mismas tasas específicas por edad que los análisis de periodo de tiempo de calendario, pero ordenadas de manera diferente. La comparación de las tasas en cohortes de nacimiento sucesivas nos permite evaluar como puede haber cambiado la incidencia de una generación a otra.

Los datos de la Tabla 4.1 pueden representarse gráficamente de diferentes maneras para ilustrar los cambios en las tasas específicas por edad frente al tiempo del calendario (tendencias seculares) o cambios de generación a generación (tendencias de cohortes).

Tabla 4.1. Mortalidad (por millón personas-año) por cáncer de pulmón en hombres, Inglaterra y Gales, 1941-78, por grupos de edad y periodo de tiempo de calendario; el año central de nacimiento se indica al final de las líneas diagonales^a

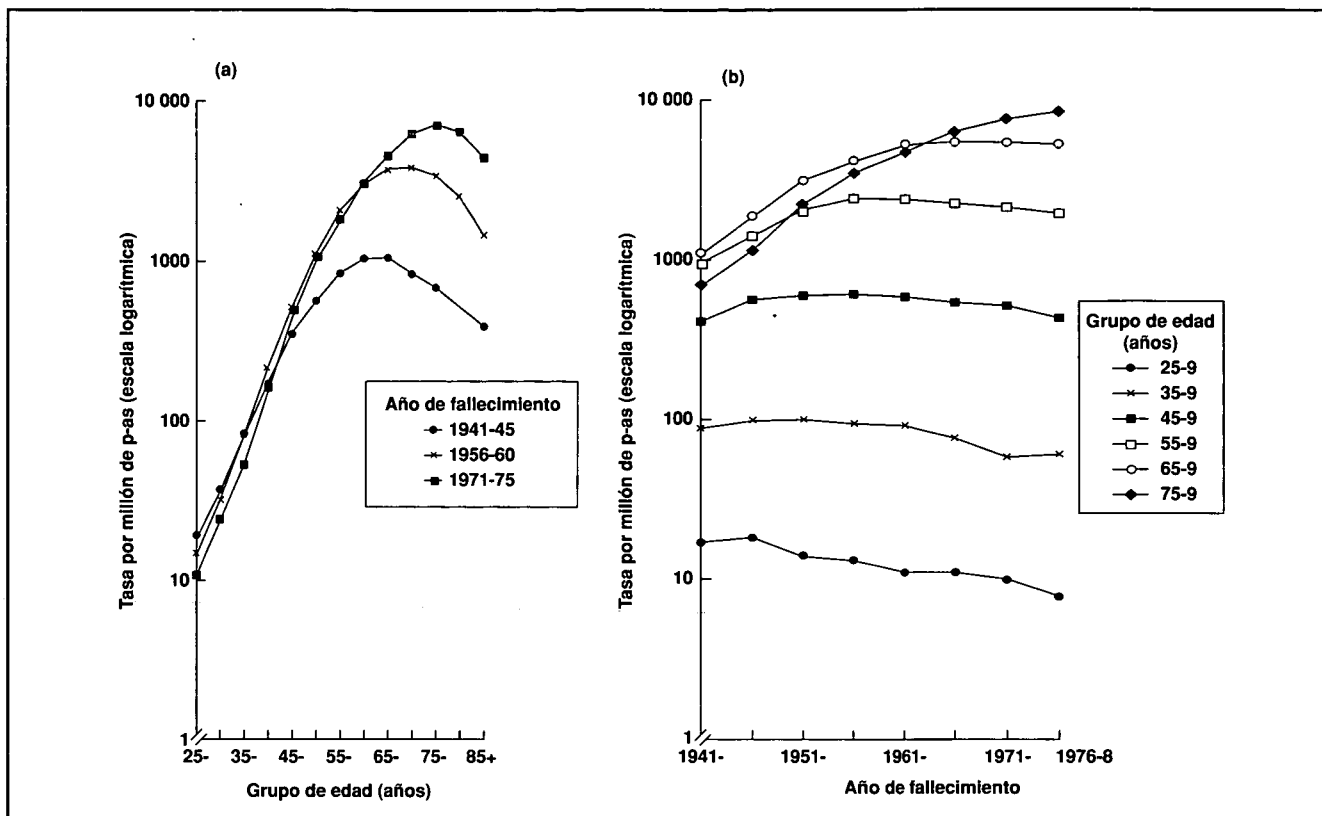


Figura 4.5. Mortalidad por cáncer de pulmón en hombres, Inglaterra y Gales, 1941-78: a) tasas presentadas para mostrar diferencias en las curvas específicas de edad entre tres periodos de calendario elegidos; b) tasas presentadas para mostrar las tendencias seculares (calendario) en las tasas específicas por edad. Para mayor claridad sólo se muestran las tasas de grupos de edad alternos; los cinco primeros grupos de edad se han omitido debido al bajo número de muertes (datos de la Tabla 4.1).

La Figura 4.5 ilustra claramente que las tendencias seculares en la mortalidad por cáncer de pulmón difieren según la edad. Las tasas aumentaron en el periodo de estudio en los grupos con edades más avanzadas pero disminuyeron en los grupos de edad más joven. Cuando las tasas se presentaban por año de nacimiento (Figura 4.6) se observaba que las tasas aumentaban para las generaciones sucesivas de hombres hasta el cambio de siglo pero disminuían para las generaciones nacidas después. Estas tendencias son prácticamente paralelas a las tendencias del hábito de fumar cigarrillos (no mostrado).

En algunas situaciones, los análisis de cohortes proporcionan una visión más comprensible de los cambios de los patrones de enfermedad ocurridos en el tiempo. Por ejemplo, si la exposición a un factor de riesgo potencial se produce muy pronto en la vida y modifica el riesgo de adquirir una enfermedad a lo largo de la vida o si en generaciones sucesivas se mantienen determinados hábitos (p.e. fumar cigarrillos y cáncer de pulmón, exposición excesiva al sol y melanoma maligno cutáneo). En otras situaciones, el análisis por intervalo de calendario puede ser más apropiado, por ejemplo, si la exposición al factor de riesgo (o factor de protección) afecta a todos los grupos de edad simultáneamente (p.e. introducción de nuevos tratamientos). No obstante, para la mayoría de las situaciones no está claro qué tipo de análisis es el más apropiado para describir las tendencias temporales y deben examinarse los resultados de ambos.

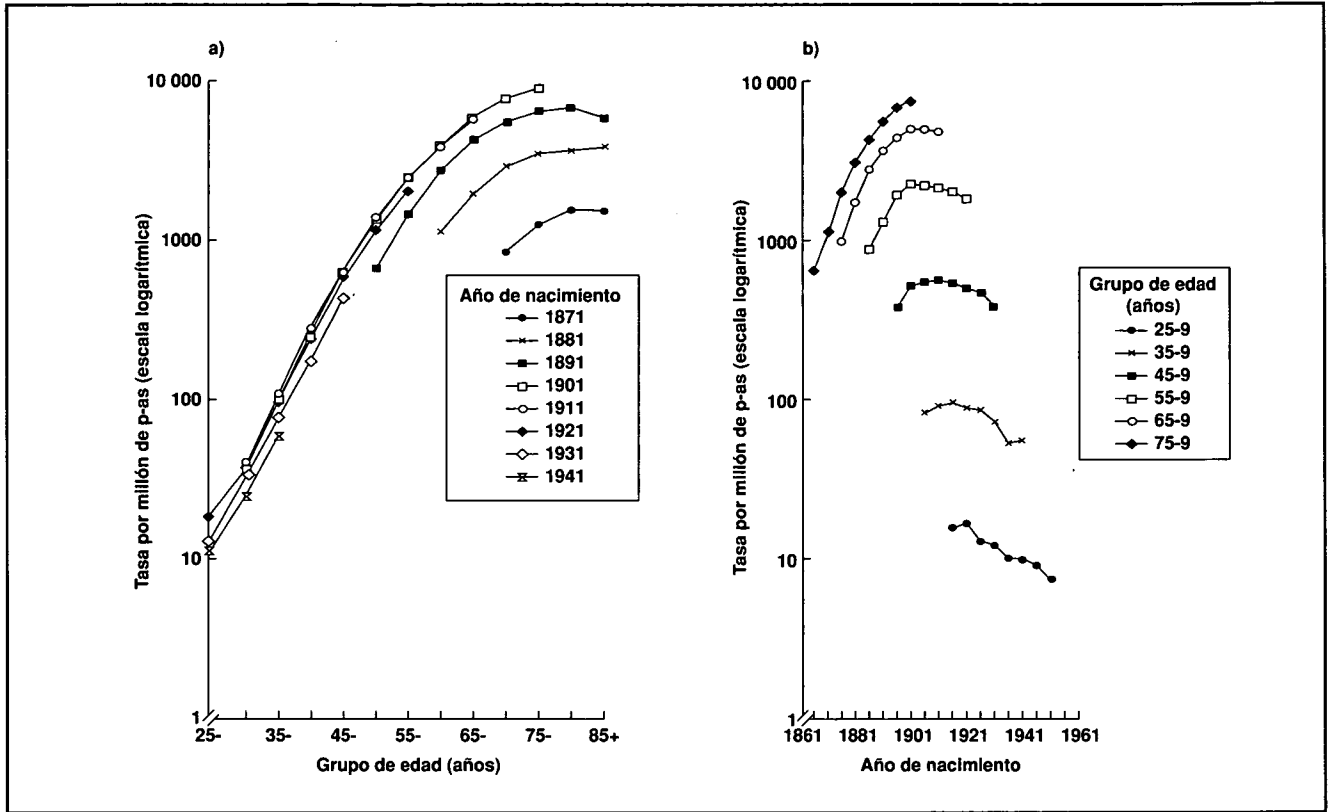


Figura 4.6. Mortalidad por cáncer de pulmón en hombres, Inglaterra y Gales, 1941-78: a) tasas presentadas para mostrar diferencias en las curvas específicas de edad para cohortes de nacimiento sucesivas; b) tasas presentadas para mostrar las tendencias en las tasas específicas por edad (para mayor claridad sólo se muestran las tasas para grupos de edad o cohortes alternos; los cinco primeros grupos de edad se omitieron debido al bajo número de muertes) (datos de la Tabla 4.1).

Los análisis descriptivos por edad, tiempo de calendario y cohorte son herramientas epidemiológicas populares para examinar los cambios temporales en la incidencia de una enfermedad. Estos análisis se basan en la inspección de tablas y gráficos de una manera similar a la mostrada en este capítulo, aunque pueden utilizarse modelos estadísticos para evaluar si existe una tendencia ascendente o descendente estadísticamente significativa en las tasas, durante el periodo de tiempo o entre cohortes de nacimiento (Clayton y Schifflers, 1987a,b).

4.3.3 Ajuste de tasas por la edad

Como se ha mencionado en el Apartado 4.3.1, las tasas crudas pueden inducir a error al comparar la incidencia entre poblaciones. Como ejemplo, comparemos la incidencia del cáncer de estómago en hombres, entre los habitantes de Cali (Colombia) y Birmingham (Inglaterra). Los datos se obtuvieron de la serie Incidencia del Cáncer en Cinco Continentes [*Cancer Incidence in Five Continents*] (Parkin et al., 1992).

Como puede observarse en la Tabla 4.2., la tasa de incidencia cruda (para todas las edades combinadas) en Birmingham fue mucho mayor que en Cali. No obstante, antes de concluir que para los hombres de Birmingham (1983-86) el riesgo de cáncer de estómago era mayor que para los de Cali (1892-86) tenemos que comparar las tasas específicas por edad de estas dos poblaciones. Sorprendentemente, éstas fueron mayores para Cali en todas las cate-

Edad (años) ^b	Cali			Birmingham		
	Nº de cánceres (1982-86)	Población masculina (1984)	Tasa anual media ^c (1982-86)	Nº de cánceres (1983-86)	Población masculina (1985)	Tasa anual media ^c (1983-86)
0-44	39	524.220	1,5	79	1.683.600	1,2
45-64	266	76.304	69,7	1037	581.500	44,6
65+	315	22.398	281,3	2352	291.100	202,0
Todas las edades	620	622.922	19,9 ^d	3.468	2.556.200	33,9

^a Datos de Parkin *et al.* (1992)

^b Para simplificar, en este ejemplo sólo se utilizan tres grupos de edad amplios.

^c Tasa por 100.000 personas-años

^d Esta tasa cruda es ligeramente menor que en el Ejemplo 4.9 (21,03 por 100.000 personas-años) debido a que los casos de edad desconocida (un total de 35) se excluyeron del presente cálculo. Sólo hubo dos casos de edad desconocida en Birmingham; su exclusión no afectó el valor de la tasa cruda.

Tabla 4.2.

Incidencia del cáncer de estómago en hombres por grupo de edad, Cali 1982-86 y Birmingham 1983-86^a.

rías de edad. La razón de esta discrepancia entre las tasas crudas y específicas por edad fue debido al hecho de que estas dos poblaciones tenían una estructura de edad considerablemente diferente (Tabla 4.3), siendo la población de Birmingham una población mucho mayor que la de Cali.

Tabla 4.3.

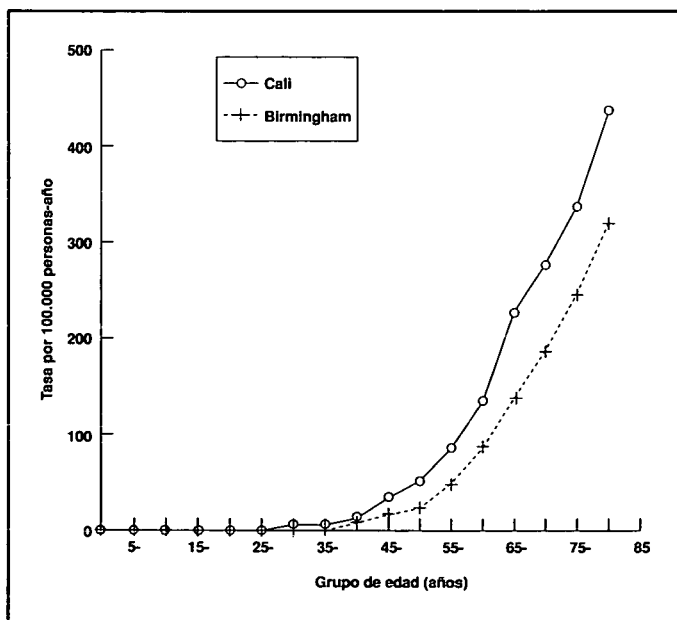
Distribución de los grupos etarios para las poblaciones masculinas de Cali, 1984 y de Birmingham, 1985^a.

Edad (años)	Porcentaje de la población masculina total	
	Cali (1984)	Birmingham (1985)
0-44	84	66
45-64	12	23
65+	4	11
Todas las edades	100	100

^a Datos de Parkin *et al.* (1992)

Figura 4.7.

Incidencia de cáncer de estómago en hombres para distintos grupos de edad en Cali, 1982-86 y en Birmingham, 1983-86 (datos de Parkin *et al.*, 1992).



El hecho de que en Cali se observe una tasa cruda inferior puede atribuirse a que su población masculina era más joven que la población masculina de Birmingham y las personas jóvenes tienen una incidencia mucho menor de cáncer de estómago que las personas mayores (Figura 4.7). Como se verá en el Capítulo 14, la edad en este ejemplo actúa como variable de confusión, es decir, está relacionada con la exposición (localidad) y es de por sí un factor de riesgo para el resultado de interés, cáncer de estómago (ver Capítulo 13).

La comparación entre las tasas de incidencia de cáncer de estómago en dos poblaciones debe prece-derse por una comparación de la estructura de edad en ambas poblaciones. Una manera de hacerlo es comparando sus tasas específicas por edad como se hizo en el ejemplo anterior. No obstante, esto puede volverse bastante engorroso al comparar varias poblaciones cada una con muchos gru-

pos de edad. Lo ideal sería tener para cada población a estudio una cifra que resumiera la experiencia de esta población de forma controlada o ajustada por las diferencias en la estructura de edad. Existen varios métodos estadísticos, algunos de los cuales se tratarán en capítulos posteriores (Capítulos 13 y 14), para controlar los efectos de las variables de confusión tales como la edad, pero en este capítulo sólo se considerará uno de ellos, la estandarización.

La estandarización es el método más común utilizado con datos de recogida sistemática. Aunque este método se emplea normalmente para ajustar el efecto de la edad, se puede utilizar para controlar cualquier otra variable de confusión como la clase social, el área de residencia, etc. Existen dos métodos de estandarización: métodos directo e indirecto.

Método directo de estandarización

Supongamos que tenemos una población hipotética, que llamaremos ‘estándar’, cuya estructura de edad se muestra en la Tabla 4.4 y supongamos que queremos saber cuántos casos de cáncer de estómago en hombres se esperarían en Cali si su población masculina tuviera la misma distribución de edad que la población estándar.

Esto es relativamente fácil de calcular como puede observarse en la Figura 4.8. Todo lo que hay que hacer es multiplicar cada tasa específica por edad de Cali por los datos de la población estándar en el grupo de edad correspondiente; la suma de todas las categorías de edad dará el número total de los casos esperados de cáncer de estómago en hombres en Cali si su población masculina tuviera la misma distribución de edad que la población estándar.

Se puede determinar cuántos casos de cáncer de estómago en hombres se esperarían en Birmingham si su población masculina tuviera la misma distribución de edad que la población estándar. Los cálculos serán iguales a los descritos anteriormente para Cali, basados en las tasas específicas por edad de Birmingham (Figura 4.8).

A partir de la suposición de que Cali y Birmingham tuvieran una estructura etaria de su población igual a la de la población estándar, pueden estimarse tasas globales de incidencia comparables dividiendo los casos totales esperados entre el total de personas-años a riesgo en la población estándar. Estas tasas se denominan tasas de inci-

Edad (años)	Población
0-44	74.000
45-64	19.000
65+	7.000
Todas las edades	100.000

Tabla 4.4. Una población estándar.

Figura 4.8. Ilustración del método directo de estandarización.

Tasas anuales medias específicas por edad en Cali, 1982-86 (por 100.000 p-as)		Tasas anuales medias específicas por edad en Birmingham, 1983-86 (por 100.000 p-as)	
Edad	Tasa	Edad	Tasa
0-44	1,5	0-44	1,2
45-64	69,7	45-64	44,6
65+	281,3	65+	202,0

Población estándar	
Edad	Población
0-44	74.000
45-64	19.000
65+	7.000

Nº de casos de cáncer de estómago en hombres esperados si la población masculina de Cali y de Birmingham tuviera la misma distribución de edad que la población estándar

a) Cali		b) Birmingham	
Edad	Casos esperados	Edad	Casos esperados
0-44	$0,000015 \times 74.000 = 1,11$	0-44	$0,000012 \times 74.000 = 0,89$
45-64	$0,000697 \times 19.000 = 13,24$	45-64	$0,000446 \times 19.000 = 8,47$
65+	$0,002813 \times 7.000 = 19,69$	65+	$0,002020 \times 7.000 = 14,14$
Total esperado	= 34,04	Total esperado	= 23,50

Tasa anual media ajustada por edad para Cali, 1982-86 =	Tasa anual media ajustada por edad para Birmingham, 1983-86 =
= $34,04/100.000 =$	= $23,5/100.000 =$
= 34,0 per 100.000 pyrs	= 23,5 per 100.000 pyrs

dencia anuales medias ajustadas por edad o estandarizadas por edad. Pero ¿cómo pueden interpretarse estas tasas? Pueden considerarse como las tasas de incidencia crudas que tendrían estas poblaciones si sus distribuciones de edad se desplazaran, de lo que eran realmente a mediados de los años 80, a la distribución de edad del estándar. Estas tasas estandarizadas son una ficción. No son las tasas de incidencia de cáncer de estómago que existían realmente en estas dos poblaciones sino las que habrían presentado si tuvieran una población hipotética en lugar de su población real, manteniendo sus propias tasas específicas por edad. No obstante, esta ficción es útil porque permite realizar comparaciones entre poblaciones de diferentes zonas o de diferentes periodos de tiempo, y estas comparaciones están exentas de la distorsión que surge de las diferencias de edad en las poblaciones existentes.

La tasa estandarizada por edad puede considerarse como una media ponderada de las tasas específicas por edad, tomando los pesos de la población estándar. Las tasas ajustadas por edad son comparables directamente siempre que se refieran a la misma población estándar, es decir, que los pesos asignados a las tasas específicas por edad sean los mismos. En el ejemplo presentado en este apartado, la tasa de incidencia anual media estandarizada por edad para Cali fue mayor que para Birmingham, lo que se corresponde con el cuadro presentado por las tasas específicas por edad. Podemos calcular una razón de tasas estandarizadas por edad dividiendo la tasa de Cali por la de Birmingham, que proporciona una razón de tasas:

$$34,0 \text{ por } 100.000 \text{ p-as} / 23,5 \text{ por } 100.000 \text{ p-as} = 1,45$$

Esta medida se denomina Razón Estandarizada de Tasas (RET) [*Standardised Rate Ratio* (SRR)] o Razón de Morbilidad ó Mortalidad Comparativa (RMA) [*Comparative Morbidity (o Mortality) Figure* (CMF)] y significa que la incidencia estimada de cáncer de estómago fue un 45% mayor en Cali que en Birmingham a mediados de los años 80 y que este exceso es independiente de las diferencias de edad entre estas dos poblaciones.

Este método de ajustar por la edad se denomina método directo de estandarización. Requiere conocer las tasas específicas por edad (o los datos para calcularlas) para todas las poblaciones en estudio y también la definición de una población estándar. La población estándar puede ser cualquier población. Puede ser una de las poblaciones que intentamos comparar o cualquier otra población. No obstante, siempre es necesario especificar qué estándar se utilizó ya que la elección de la población estándar puede afectar la comparación. Se han creado poblaciones estándar 'convencionales', tales como las poblaciones estándar Mundial y estándar Europeo y tienden a utilizarse ampliamente para realizar las comparaciones directas de las tasas (véase Anexo 4.1). De hecho, la población estándar presentada en la Tabla 4.4 es un resumen de la población mundial estándar.

Para simplificar, en el ejemplo presentado en este apartado sólo se han utilizado tres grupos de edad amplios. No obstante, esto no proporciona un

ajuste de edad adecuado y deben utilizarse bandas de edad más refinadas. Es tradicional utilizar grupos de edad de 5 años porque es el corte de edad utilizado con mayor frecuencia en las publicaciones de datos específicos de cáncer. Si en nuestro ejemplo de cáncer de estómago (Figura 4.8) se hubiera realizado el ajuste de edad en grupos de edad de 5 años, las tasas ajustadas de Cali y Birmingham habrían sido 36,3 y 21,2 por 100.000 personas-años, respectivamente, la razón de tasas sería 1,71. Cuando las tasas cambian mucho con la edad, pueden ser necesarios unos cortes de edad incluso más detallados (p.e. cada año de vida) para obtener un ajuste de edad adecuado.

Cabe recordar que una tasa estandarizada por edad es una tasa artificial y no una tasa real. Permite comparar la incidencia de una enfermedad en una población con la de otra población, con la edad controlada. No debe usarse si lo que se quiere es sólo una determinación exacta de la aparición de la enfermedad en una población en lugar de una comparación entre poblaciones.

Método indirecto de estandarización

Supongamos ahora que se conoce el número total de casos de cáncer de estómago en Cali en 1982-86 pero no se dispone de su distribución (Tabla 4.5). En estas circunstancias, no puede utilizarse el método directo de estandarización.

Tabla 4.5.
Incidencia del cáncer de estómago en Cali, 1982-86 y Birmingham, 1983-86^a.

Edad (años) ^b	Cali			Birmingham		
	Nº de cánceres (1982-86)	Población masculina (1984)	Tasa anual media ^b (1982-86)	Nº de cánceres (1983-86)	Población masculina (1985)	Tasa anual media ^b (1983-86)
0-44	N/D	524.220	-	79	1.683.600	1,2
45-64	N/D	76.304	-	1.037	581.500	44,6
65+	N/D	22.398	-	2.352	291.100	202,0
Todas las edades	620	622.922	19,9	3.468	2.556.200	33,9

N/D = datos supuestamente no disponibles (ver Tabla 4.2.).
^a Datos de Parkin *et al.*, 1992.
^b Tasa por 100.000 p-as.

No obstante, pueden calcularse cuántos casos de cáncer de estómago en hombres se esperarían en Cali y Birmingham si los hombres de estas dos poblaciones presentaran las mismas tasas de incidencia específicas por edad que en Birmingham. En otras palabras, las tasas específicas por edad de Birmingham se tomarán como una serie de tasas estándar. Los cálculos de presentan en la Figura 4.9. El número esperado de casos de cáncer en Cali se calcula multiplicando las tasas específicas por edad anuales medias de Birmingham por personas-años a riesgo de Cali en el grupo de edad correspondiente; la suma de las categorías de edad dará el número total de casos de cáncer en hombres que se esperarían en Cali si su población masculina tuviera las mismas tasas de incidencia específicas por edad para el cáncer de estómago que la población masculina de Birmingham. Obvia-

Tasas anuales medias específicas por edad de Birmingham, 1983-86 (por 100.000 p-as)			
Edad		Tasa	
0-44		1,2	
45-64		44,6	
65+		202,0	

Total personas-años a riesgo en Cali, 1982-86			Total personas-años a riesgo en Birmingham, 1983-86		
Edad	Personas-años		Edad	Personas-años	
0-44	$524.220 \times 5 = 2.621.100$		0-44	$1.683.600 \times 4 = 6.734.400$	
45-64	$76.304 \times 5 = 381.520$		45-64	$581.500 \times 4 = 2.326.000$	
65+	$22.398 \times 5 = 111.990$		65+	$291.100 \times 4 = 1.164.400$	
Total	= 3.114.610		Total	= 10.224.800	

Número de casos de cáncer de estómago en el hombre, si las poblaciones tuvieran las mismas tasas de incidencia específicas por edad de cáncer de estómago que Birmingham

a) en Cali		b) en Birmingham	
Edad	Casos esperados	Edad	Casos esperados
0-44	$0,000012 \times 2.621.100 = 31,45$	0-44	$0,000012 \times 6.734.400 = 79$
45-64	$0,000446 \times 381.520 = 170,15$	45-64	$0,000446 \times 2.326.400 = 1.037$
65+	$0,002020 \times 111.990 = 226,22$	65+	$0,002020 \times 1.164.400 = 2.352$
Total esperado (E), 1982-86	427,82	Total esperado (E), 1983-86 =	3.468
Total observado (O), 1982-86	620	Total observado (O), 1983-86 =	3.468
O/E (%) =	145	O/E (%) =,	100

Figura 4.9. Método indirecto de estandarización.

mente, el número de casos esperados en Birmingham es igual al número observado.

Estos casos esperados de cáncer de estómago se refieren a lo que pasaría si las dos localidades tuvieran las mismas tasas de incidencia específicas por edad de cáncer de estómago y no la misma estructura de la población. Así no tendría sentido calcular unas tasas globales para cada localidad dividiendo el número total de casos esperados por el correspondiente personas-años a riesgo para esta localidad. Pero dentro de cada localidad pueden compararse los números de casos observados y esperados porque ambos se refieren a la misma población. La relación entre el número observado y esperado de casos se denomina Razón de Incidencia Estandarizada (RIE) [Standardised Incidence Ratio (SIR)] o Razón de Mortalidad Estandarizada (RME) [Standardised Mortality Ratio (SMR)] si un caso se define como muerte y generalmente se expresa como un porcentaje.

En el ejemplo anterior, la RIE (%) para Birmingham es igual a 100 lo que significa que el número de casos observados es igual al número de casos esperados utilizando como tasas estándar las tasas de incidencia específicas de edad de cáncer de estómago para Birmingham. La RIE (%) para Cali es igual a 145 lo que significa que el número de casos observados fue el 45% superior que el número esperado si Cali tuviera la misma incidencia de cáncer de estómago que Birmingham. Este resultado es similar al obtenido previamente con el método directo.

Este método se denomina método indirecto de estandarización. Al igual que con el método directo, los resultados dependen en parte del estándar elegido. No obstante, el método indirecto es menos sensible a la elección de la población estándar que el método directo.

¿Cuál es el mejor método?

¿Cuál es el mejor método?

Al comparar la aparición de enfermedad o mortalidad en dos o más poblaciones, la estandarización directa e indirecta tienden a proporcionar unos resultados bastante similares en la práctica. No obstante, en la elección del método pueden influir varias consideraciones:

- (1) El método directo requiere tasas específicas por estrato (p.e. tasas específicas por edad) en todas las poblaciones en estudio. El método

indirecto sólo requiere el número total de casos que se produjeron en cada población en estudio, no su división en estratos. Si no se dispone de las tasas específicas por estrato para la población en estudio, el método indirecto puede ser la única estrategia posible.

- (2) El ajuste indirecto es preferible al método directo cuando las tasas específicas por edad se basan en números bajos de sujetos. Las tasas utilizadas en el ajuste directo estarían basadas en estos números bajos y por tanto estarían sometidas a una variación sustancial del muestreo (véase Apartado 6.1.4). Con el método indirecto podemos elegir las tasas más estables como estándar para asegurar que las tasas globales sean lo más precisas posible.
- (3) En general, la comparación de la incidencia en dos o más poblaciones mediante la estandarización directa está sometida a un error menor que mediante la estandarización indirecta. Las razones escapan al marco de este texto y no se comentarán aquí.

En Breslow y Day (1987), páginas 72-76 se presenta una discusión más detallada de las ventajas e inconvenientes de cada método de estandarización.

¿Es siempre adecuado utilizar estas medidas globales 'ajustadas'?

Estas medidas ajustadas por edad son un resumen útil de las tasas específicas por edad, aunque las propias tasas específicas por edad proporcionan la mayor parte de la información. Cabe destacar que, en algunas circunstancias, puede no ser apropiado resumir las tasas de enfermedad en una medida global única. A continuación se presenta un ejemplo de cómo pueden ser engañosas las tasas estandarizadas por edad. Supongamos que estamos interesados en monitorizar las tendencias en la mortalidad por cáncer de ovario en Inglaterra y Gales. Las tasas de mortalidad ajustadas por edad para este cáncer en Inglaterra y Gales aumentaron ligeramente de 1970-74 a 1985-89, como se muestra en la Figura 4.10a. No obstante, las tendencias en las tasas específicas por edad para este periodo mostraron que el aumento no se produjo en todos los grupos de edad. La Figura 4.10b muestra la razón de las tasas específicas por edad de 1985-89 a 1970-74. Se observa que las tasas de mortalidad por este cáncer aumentaban en las edades mayores (razón de tasas mayor de 1), pero no en edades menores de 55 años (razón de tasas menor de 1). Esta información se perdió cuando se consideraron las tasas de mortalidad estandarizadas por edad para todas las edades combinadas, debido a que las tasas de mortalidad por este cáncer en edades más jóvenes eran tan bajas que estaban dominadas por las tasas de mortalidad mucho más altas en los grupos de edad mayores.

Con objeto de evaluar si es apropiado utilizar medidas globales ajustadas, siempre deben examinarse antes las relaciones de tasas específicas por edad. Si existe una evidencia de que estas relaciones varían sistemáticamente con la edad, esta información se perdería inevitablemente en la medida global ajustada por edad.

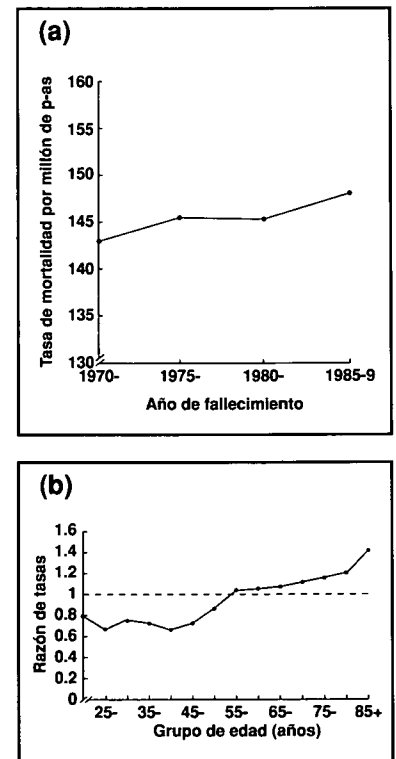


Figura 4.10. Mortalidad por cáncer de ovario en Inglaterra y Gales, 1970-89. (a) Tasas estandarizadas por edad en la población femenina de Inglaterra y Gales en 1981; (b) Razones de tasas de mortalidad específicas por edad, 1970-1974 y 1985-1989 (las tasas en 1970-1974 se consideran el valor basal).

4.3.4 Tasa acumulada

La tasa acumulada es otra medida de la aparición de la enfermedad cada vez más utilizada en epidemiología del cáncer y se ha incluido en los últimos volúmenes de la serie Incidencia del Cáncer en Cinco Continentes [*Cancer Incidence in Five Continents*] (Parkin *et al.*, 1992).

Una tasa acumulada es igual a la suma de las tasas de incidencia específica por edad, en un rango de edad determinado. Se debe especificar el periodo de edad sobre el cual se acumula la tasa y dependerá de la comparación a realizar. Por ejemplo, para tumores infantiles, se tomaría la edad de 0-14. No obstante, en general, el riesgo durante toda la vida sería la medida adecuada, que puede considerarse como 0-74.

La tasa acumulada puede calcularse sumando las tasas de incidencia específicas por edad (siempre que se expresen en las mismas unidades persona-tiempo – p.e. 100.000 personas-años) multiplicado por la amplitud del grupo de edad. Así, para grupos de edad de cinco años, la tasa acumulada sería cinco veces la suma de las tasas de incidencia específica por edad en el intervalo de edad en cuestión (Figura 4.11). Si los grupos de edad tienen una amplitud diferente, cada tasa específica por edad se multiplicaría primero por la amplitud del grupo de edad; la suma de todas las categorías de edad proporcionaría la tasa acumulada.

La tasa acumulada puede interpretarse como una forma de estandarización directa por edad con el mismo tamaño de población (es decir, denominador) en cada grupo de edad. Por tanto, evita la arbitrariedad implica-

Figura 4.11. Cálculo de la tasa acumulada en el intervalo de edad de 0-74 años y el riesgo acumulado de cáncer de estómago masculino, Birmingham, 1983-86 (datos de Parkin *et al.*, 1992).

Grupo de edad (años)	Tasa de incidencia específica por edad (por 100.000 p-as)
0-4	0
5-9	0
10-14	0
15-19	0
20-24	0,1
25-29	0,1
30-34	0,9
35-39	3,5
40-44	6,7
45-49	14,5
50-54	26,8
55-59	52,6
60-64	87,2
65-69	141,7
70-74	190,8
Total	524,9
Total × 5	2.624,5
Tasa acumulada = 100 × (2.624,5/100.000) = 2,6%	
Riesgo acumulado = 100 × {1-exp(-tasa acumulada/100)} = 2,6%	

da en la elección de una población estándar. Esta medida se suele expresar como porcentaje.

Otra ventaja de la tasa acumulada es que proporciona una estimación del riesgo acumulado, es decir, el riesgo que tendría un individuo de desarrollar un cáncer en particular en un periodo de vida definido si no actuaran otras causas de muerte. El valor del riesgo acumulado puede calcularse utilizando la fórmula siguiente:

$$\text{Riesgo acumulado} = 100 \times \{1 - \exp(- \text{tasa acumulada} / 100)\}$$

No obstante, si el valor de la tasa acumulada es menor del 10%, su valor es prácticamente igual al del riesgo acumulado. Así, en la Figura 4.11, el riesgo estimado de que un hombre de Birmingham desarrolle cáncer de estómago entre las edades de 0-74 años es del 2,6% (suponiendo que no actúan otras causas de muerte). Este valor es igual al valor de la tasa acumulada.

La Tabla 4.6 muestra los valores de las tasas crudas, las tasas estandarizadas por edad para 5 años y las tasas acumuladas para el cáncer de estómago en hombres en Cali y Birmingham. A diferencia de las tasas crudas, las tasas estandarizadas por edad y las tasas acumuladas proporcionan un cuadro correcto de la incidencia relativa de cáncer de estómago en las dos poblaciones.

	Cali, 1982-86	Birmingham, 1983-86	Razón de tasas
Tasa cruda (por 100.000 p-as)	19,9	33,9	0,59
Tasa estandarizada por edad ^a (por 100.000 p-as)	36,3	21,2	1,71
Tasa acumulada, 0-74 años (%)	4,6	2,6	1,77

^a Tasas estandarizadas a la población estándar mundial. Las tasas estandarizadas por edad difieren ligeramente de las de la Figura 4.8 porque las tasas mostradas aquí se han ajustado por grupos de edad de 5 años.

Tabla 4.6.

Tasas de incidencia crudas (promedio anual), tasas estandarizadas por edad (promedio anual) y tasas acumuladas para el cáncer de estómago en hombres, Cali, 1982-86 y Birmingham, 1983-86.

4.3.5 Falta de denominadores adecuados

En algunas situaciones no existen denominadores adecuados que nos permitan calcular una de las medidas de incidencia comentadas hasta ahora. Esto puede producirse bien porque no existan datos del denominador (p.e. no se ha efectuado un censo), o porque no pueda definirse una población de referencia (p.e. para un registro hospitalario), o porque la búsqueda de casos haya sido lo suficientemente incompleta como para que los denominadores derivados de otras fuentes (p.e. el censo) no sean comparables con los datos del numerador. En estas circunstancias, tradicionalmente se han calculado medidas proporcionales, esto es, se ha expresado el número de casos de una enfermedad particular como una proporción del número total de casos:

$$\text{Incidencia proporcional (\%)} = \frac{\text{N}^\circ \text{ de casos de una enfermedad determinada en un período de tiempo especificado}}{\text{Número total de casos (de todas las enfermedades) en el mismo período de tiempo}} \times 100$$

Las comparaciones entre las poblaciones pueden efectuarse entonces por cálculo de las Razones de Incidencia (o Mortalidad) Proporcional (RIP ó RMP) [*Proportional Incidence (o Mortality) Ratios* (PIR ó PMR)]. Estas razones se calculan como sigue:

$$\text{RMP (\%)} = \frac{\text{Proporción de casos de una enfermedad específica en la población A}}{\text{Proporción de casos de la misma enfermedad en la población B}} \times 100$$

Al igual que las tasas, estas proporciones también pueden estandarizarse por la edad (o cualquier otro factor de confusión potencial).

Una medida proporcional no es equivalente a una tasa, ya que el denominador se obtiene del número total de casos, no de la población a riesgo. Las medidas proporcionales responden a la pregunta «¿Qué proporción de casos (o muertes) es atribuible a una enfermedad en particular?» Por el contrario, una tasa específica de la causa responde a la pregunta «¿Cuál es el riesgo de desarrollar (o de morir por) una enfermedad en particular para los miembros de una población determinada?»

Las medidas proporcionales pueden ser engañosas debido a que su denominador se refiere al número total de casos (o de muertes), que depende del número de casos (o muertes) por otras causas aparte de la enfermedad considerada. Por ejemplo, aunque la proporción de muertes debidas a cáncer es mayor en mujeres jóvenes que en mujeres mayores, las tasas de mortalidad en realidad son mayores entre las mujeres de mayor edad (Figura 4.12). La explicación es, por supuesto, que el número total de muertes por otras causas, especialmente por enfermedades cardiovasculares, también es considerablemente mayor en mujeres mayores. Por tanto, aunque el número total de muertes por cáncer sea mayor en mujeres mayores, éstas constituyen una proporción menor del total de muertes que a edades medias.

Estas medidas (y el uso de *odds ratio* como alternativa) se seguirán comentando en el Capítulo 11.

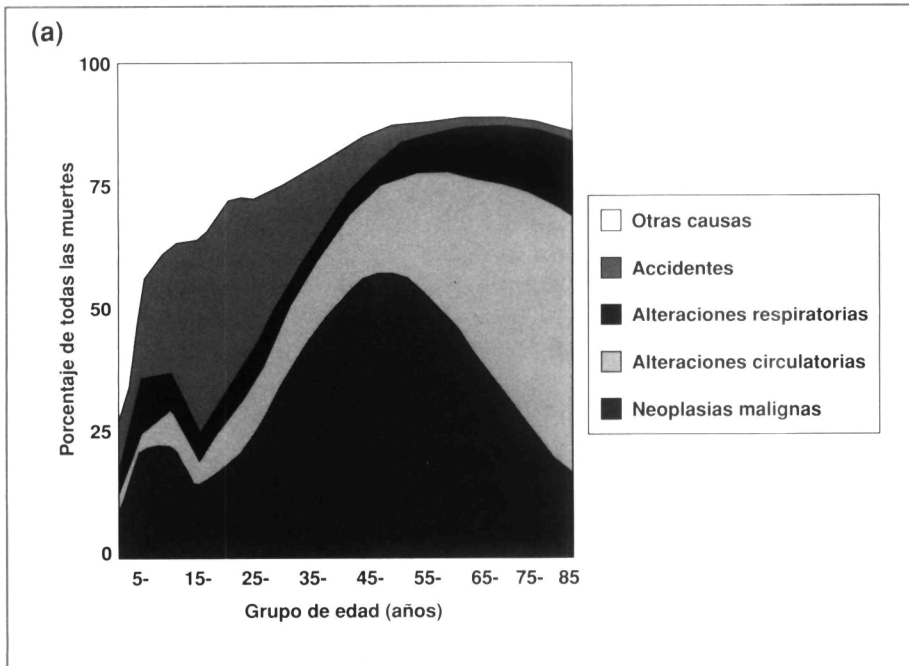
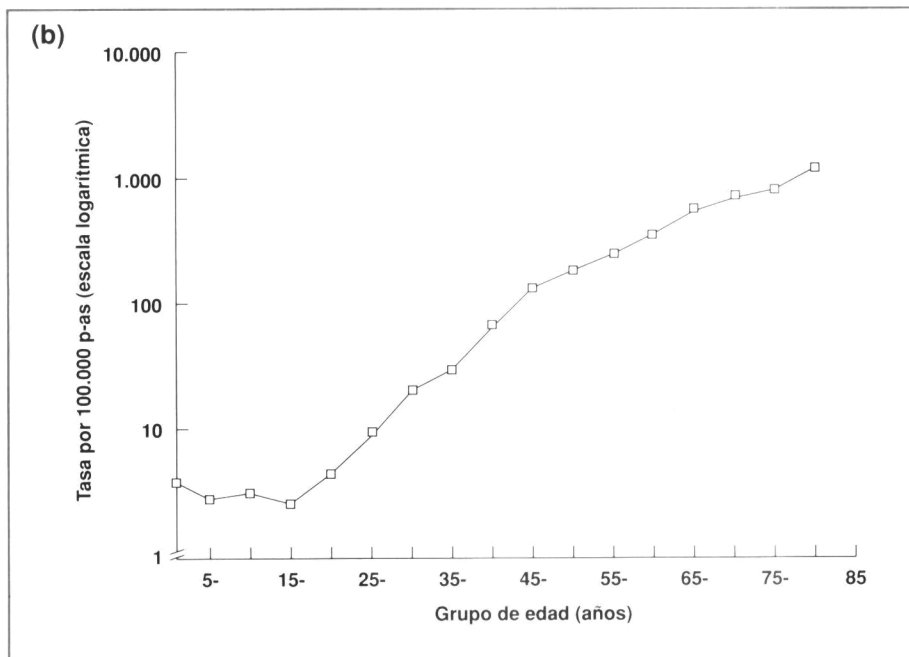


Figura 4.12. Muertes femeninas en Inglaterra y Gales, 1993. (a) Proporción de muertes debidas a cáncer y otras causas, por edad; (b) tasas de mortalidad por cáncer en mujeres, por edad (datos de OPCS, 1995).



Lecturas adicionales

* La mayoría de las medidas de aparición de la enfermedad comentadas aquí se tratan con más detalle en Breslow y Day (1987) y Estève *et al.* (1994).

* En Clayton y Schiffers (1987a,b) se presenta una discusión más elaborada de los efectos de la edad, tiempo de calendario y cohorte.

Cuadro 4.1. Puntos esenciales

- La cuantificación de enfermedad y otros acontecimientos relacionados con la salud requiere una definición clara de los casos (numerador), la población de riesgo (denominador) y el intervalo de tiempo a los que se refieren los datos del numerador y del denominador.
- Existen dos medidas principales de la aparición de enfermedad: la prevalencia y la incidencia. La *prevalencia* se refiere al número total de casos existentes (anteriores y nuevos) de una enfermedad particular en una población en un momento determinado en el tiempo. La *incidencia* se refiere a la aparición de casos nuevos en una población durante un periodo de tiempo especificado.
- La incidencia puede medirse como *riesgo*, *odds de la enfermedad* o *tasa*. El cálculo del riesgo requiere el seguimiento completo de todos los sujetos en estudio durante todo el periodo del estudio. Por el contrario, el cálculo de las tasas considera las diferencias entre las duraciones individuales del seguimiento.
- Las medidas de la aparición de enfermedad pueden calcularse bien como medidas crudas para toda la población, bien como medidas específicas de estrato, es decir, de modo separado para ciertos subgrupos de la población.
- La incidencia de una enfermedad varía con el tiempo. Estos cambios se producen simultáneamente en tres escalas de tiempo diferentes: edad, periodo del calendario (tendencias seculares) y fecha de nacimiento (tendencias de cohortes) pero pueden examinarse por separado en un diagrama de dos ejes, edad y periodo de tiempo de calendario (diagrama de Lexis). Las diagonales en este diagrama representan cohortes de nacimiento sucesivas.
- Las tasas crudas pueden ser engañosas al comparar la incidencia de diferentes poblaciones ya que no tienen en cuenta las diferencias en la estructura de edad. La *estandarización por edad* (directa e indirecta) puede utilizarse para obtener medidas resumen 'ajustadas' según las diferencias en la estructura de edad. Como alternativa pueden calcularse las *tasas acumuladas*.
- Cuando no existan denominadores adecuados pueden calcularse medidas proporcionales tales como *razones de mortalidad* (o *incidencia*) *proporcional*. Éstas deben interpretarse con precaución ya que su denominador es el número total de casos y no la población de riesgo.

Anexo 4.1

Poblaciones Estándar Convencionales

La elección de la población estándar a utilizar en el método directo de estandarización es, en cierto grado, arbitraria. Por ejemplo, si el problema fuera comparar la aparición de la enfermedad en varios grupos en Inglaterra y Gales, un estándar adecuado podría ser la población adulta en Inglaterra y Gales. Por otro lado, no existe ninguna buena razón para elegir la población adulta en Inglaterra y Gales como estándar si se está interesado en realizar comparaciones internacionales.

Para comparaciones internacionales se han utilizado diferentes poblaciones estándar convencionales (Tabla a4.1). Estas poblaciones estándar varían desde la población 'Africana', con una baja proporción de personas mayores, pasando por una población 'Mundial' intermedia, a una población 'Europea', con una elevada proporción de personas mayores. En los primeros volúmenes de la serie *Incidencia del Cáncer en Cinco Continentes* [*Cancer Incidence in Five Continents*] las tasas estaban estandarizadas a estas tres poblaciones, pero los estándares 'Europeo' y 'Africano' se eliminaron en el volumen IV (Waterhouse *et al.*, 1982) y se sustituyeron por las tasas acumuladas en los intervalos de edad 0-64 y 0-74 años.

La población truncada es la población 'Mundial' limitada a los grupos de edad media. Ésta se utilizó con frecuencia en el pasado porque era más probable que los datos para los grupos de mayor edad fueran menos fiables que los datos de los grupos de edad media y porque para la mayoría de las formas de cáncer virtualmente no existen casos en edades inferiores a 35 años.

Tabla A.4.1.
Poblaciones estándar convencionales
utilizadas en comparaciones internacio-
nales.

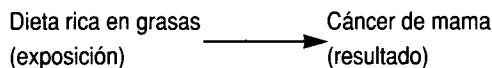
Grupo de edad (años)	Africano	Mundial	Europeo	Truncado
0	2.000	2.400	1.600	—
1-4	8.000	9.600	6.400	—
5-9	10.000	10.000	7.000	—
10-14	10.000	9.000	7.000	—
15-19	10.000	9.000	7.000	—
20-24	10.000	8.000	7.000	—
25-29	10.000	8.000	7.000	—
30-34	10.000	6.000	7.000	—
35-39	10.000	6.000	7.000	6.000
40-44	5.000	6.000	7.000	6.000
45-49	5.000	6.000	7.000	6.000
50-54	3.000	5.000	7.000	5.000
55-59	2.000	4.000	6.000	4.000
60-64	2.000	4.000	5.000	4.000
65-69	1.000	3.000	4.000	—
70-74	1.000	2.000	3.000	—
75-79	500	1.000	2.000	—
80-84	300	500	1.000	—
85+	200	500	1.000	—
Total	100.000	100.000	100.000	31.000

Capítulo 5

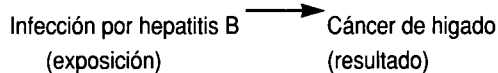
Revisión de los diseños de estudios

Cuantificar la aparición de una enfermedad u otro suceso relacionado con la salud es sólo el inicio del proceso epidemiológico. Los epidemiólogos también están interesados en evaluar si una exposición está asociada a una enfermedad determinada en una población (o cualquier otro resultado de interés). Por ejemplo, los investigadores pueden estar interesados en obtener una respuesta a preguntas como las siguientes:

- * *¿Aumenta el riesgo de cáncer de mama una dieta rica en grasas?*



- * *¿Aumenta el riesgo de cáncer de hígado la infección por el virus de la hepatitis B?*



El primer paso en un estudio epidemiológico es definir la hipótesis a analizar. Ésta debe incluir una definición precisa de la(s) exposición(es) y del(os) resultado(s) a estudio. El siguiente paso es decidir qué diseño de estudio será el más apropiado para probar la hipótesis específica del estudio.

5.1 Tipos de diseño de estudio

Básicamente existen dos métodos para evaluar si una exposición está asociada a un resultado en particular: experimental y observacional. El método experimental es probablemente el más familiar para los clínicos, ya que corresponde a la metodología utilizada en los estudios de investigación realizados en el laboratorio. En un experimento, los investigadores estudian el impacto de la variación de algún factor que pueden controlar. Por ejemplo, los investigadores pueden estudiar una camada de ratas, elegir al azar la mitad para exponerlas a un supuesto agente carcinogénico y después determinar la frecuencia con que se desarrolla el cáncer en los dos grupos. El equivalente de este experimento de animales en seres humanos implicaría la selección de un grupo de individuos, la exposición aleatoria de la mitad de ellos a un hipotético factor productor de enfermedad y la comparación de la aparición de la enfermedad entre los expuestos y los no expuestos. Por razones éticas, sería imposible realizar un estudio así en humanos. No obstante, puede realizarse un ensayo para evaluar si la eliminación de la exposición disminuye la incidencia y la mortalidad. Por lo tanto,

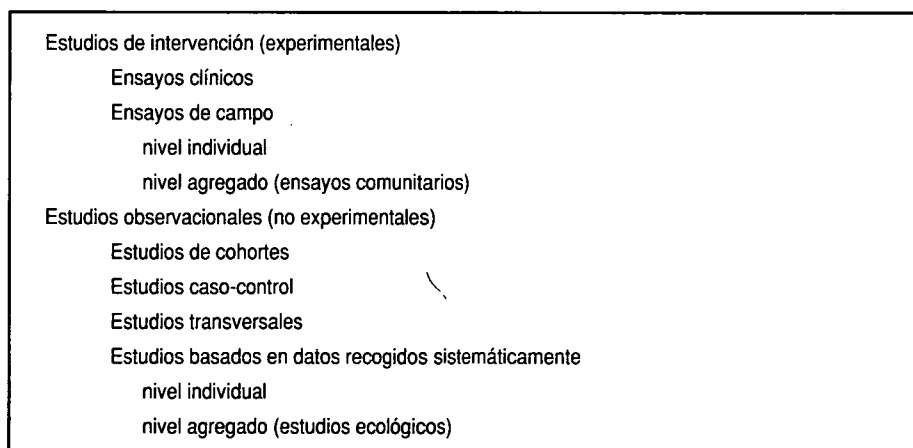
los estudios experimentales en epidemiología están limitados a intervenciones de las que se espera un beneficio potencial.

Debido a que los epidemiólogos raramente pueden realizar experimentos, su papel está limitado en gran parte a observar la aparición de la enfermedad en personas que ya pertenecen a grupos diferentes. Por ejemplo, podemos seguir a personas infectadas o no con el virus de la hepatitis B y comprobar si el riesgo de padecer cáncer de hígado es diferente. Estos estudios se denominan observacionales porque el papel del investigador es meramente ser observador de lo que pasa, anotando quien está o no expuesto y quien ha desarrollado el resultado de interés y quien no.

Un problema importante en los estudios observacionales es que los grupos observados pueden diferir en muchas otras características, además de la estudiada. Así, las personas con diferentes ocupaciones pueden diferir no sólo en la exposición a los riesgos ocupacionales sino también en otras características del estilo de vida tales como el nivel socioeconómico, el estado de salud, la actividad física en el trabajo, el hábito tabáquico y de alcohol y otros muchos factores. Debido a estos factores de confusión, que muchas veces no son cuantificables, establecer el papel de una exposición en un estudio observacional es más difícil que en un estudio experimental.

Por lo tanto, los estudios epidemiológicos en general pueden clasificarse en estudios de intervención y estudios observacionales. Dentro de cada una de estas dos amplias categorías, definimos los estudios que se muestran en la Figura 5.1.

Figura 5.1.
Principales tipos de diseño de los estudios epidemiológicos



El esquema de clasificación mostrado en la Figura 5.1 se debe considerar sólo como una manera simple de presentar y comentar los diferentes diseños de estudios. Por supuesto, existen muchas otras maneras de clasificarlos. En realidad, muchos estudios pueden presentar características de más de un diseño.

5.1.1 Estudios de intervención

Los estudios de intervención se caracterizan en que la distribución de los sujetos del estudio en los diferentes grupos la realiza el investigador me-

dante la *aleatorización* (Figura 5.2). Este método asegura que la asignación de los sujetos a los diferentes grupos se determina sólo por el azar y no está influenciada de manera subjetiva por el investigador o los participantes (Apartado 7.9.1). Existen dos tipos principales de estudios de intervención: ensayos clínicos y ensayos de campo.

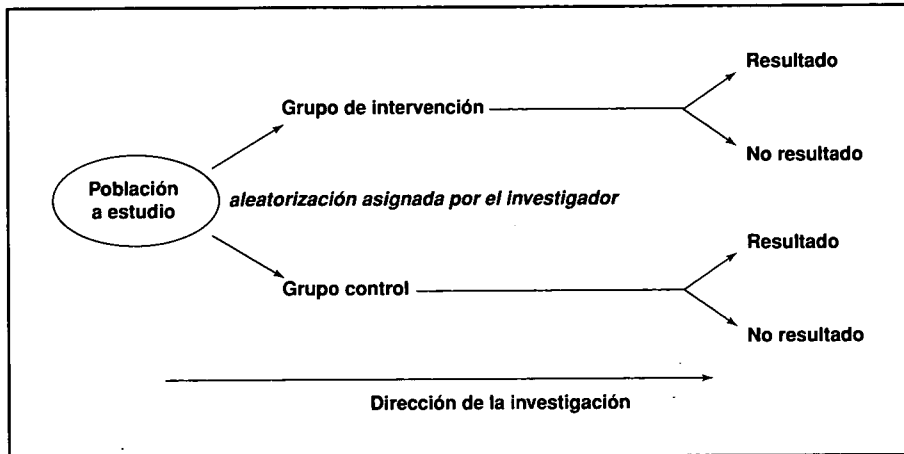


Figura 5.2.
Perfil de un estudio de intervención

El objetivo de un ensayo clínico es evaluar nuevas formas de tratamiento de una enfermedad u otro evento de interés. Así, los ensayos clínicos normalmente se efectúan en hospitales o clínicas, en personas que ya han desarrollado la enfermedad.

Ejemplo 5.1. *El Estudio de Intervención sobre Nutrición en la Mujer [Women's Intervention Nutrition Study] se realizó para evaluar si una dieta pobre en grasas reduciría la recidiva de cáncer y mejoraría la supervivencia en mujeres intervenidas de cáncer de mama en estadios iniciales e intermedios. Se incluyeron un total de 2.000 mujeres postmenopáusicas intervenidas de cáncer de mama que se distribuyeron aleatoriamente en dos grupos. Un grupo recibió el tratamiento estándar más una dieta pobre en grasas, y el otro sólo el tratamiento estándar. El tratamiento estándar consiste en tamoxifeno y radioterapia y/o quimioterapia según proceda. El estudio durará 5 años (Chlebowski y Grosvenor, 1994; Henderson, 1995).*

En el Ejemplo 5.1, las mujeres que ya habían desarrollado el cáncer de mama (y habían sido operadas) se distribuyeron aleatoriamente en tratamiento estándar acompañado de un régimen dietético pobre en grasas ('grupo de intervención') o sólo tratamiento estándar ('grupo control') para evaluar si una dieta pobre en grasa reduce la recidiva del tumor y aumenta la supervivencia.

A diferencia de los ensayos clínicos, los ensayos de campo tratan con sujetos que no tienen la enfermedad y por tanto generalmente se realizan en una comunidad en lugar de hospitales o clínicas. Su principal objetivo es

evaluar si un agente o procedimiento reduce el riesgo de desarrollar la enfermedad entre los que no la tienen en el momento de la inclusión.

Ejemplo 5.2. *La Iniciativa de Salud de la Mujer [Women's Health Initiative] es un ensayo aleatorizado en curso en los EE.UU. para determinar si una dieta pobre en grasas reduce la incidencia de cáncer de mama. Un total de 48.000 mujeres postmenopáusicas sin historia previa de cáncer de mama o de colon se distribuyen aleatoriamente en el grupo de intervención o de control. La intervención en la dieta está destinada a reducir la ingesta de grasa en un 20% de las kilocalorías totales y aumentar la ingesta de frutas y vegetales. El ensayo durará 11 años (Chlebowski y Grosvenor, 1994; Henderson, 1995).*

En el Ejemplo 5.2, las mujeres se distribuyen aleatoriamente para recibir ('grupo de intervención') o no recibir ('grupo control') el programa de intervención sobre la dieta.

Mientras que las complicaciones del cáncer de mama son bastante comunes en mujeres que ya han padecido la enfermedad, el riesgo en mujeres sanas de desarrollar cáncer de mama es muy bajo, incluso en poblaciones de alto riesgo. En consecuencia, el ensayo de campo del Ejemplo 5.2 requiere un número de sujetos mucho mayor y seguidos durante un periodo de tiempo mucho más largo que el ensayo clínico del Ejemplo 5.1.

Ejemplo 5.3. *El Estudio de Intervención sobre la Hepatitis en Gambia [The Gambia Hepatitis Intervention Study] es un proyecto de vacunación a gran escala contra la hepatitis B iniciado en Gambia en julio de 1986. En este ensayo, 60.000 niños recibieron la vacuna de la hepatitis B y un número similar no la recibieron. Un sistema nacional de vigilancia detectará todos los casos nuevos de carcinoma hepatocelular (y otras enfermedades hepáticas crónicas) durante un periodo de 30 a 40 años (The Gambia Hepatitis Study Group, 1987).*

En el Ejemplo 5.3, la mitad de los niños que participan en el estudio recibieron la vacuna de la hepatitis B ('grupo de intervención') y la otra mitad no ('grupo control'). Este estudio tiene un periodo de seguimiento notablemente alto (¡no se podrá responder a la principal pregunta de la investigación antes del año 2020!). Esto es debido a que la intervención debía ser en el primer año de vida antes de una posible infección por el virus de la hepatitis B, mientras que la mayoría de los casos de carcinoma hepatocelular se esperan en la edad adulta.

Estos dos últimos ejemplos ilustran claramente la complejidad del diseño y la realización de los ensayos de campo cuando el cáncer es el resultado de interés. Debido a que el cáncer es una enfermedad relativamente rara, estos estudios requieren la inclusión de grandes cantidades de sujetos y un seguimiento durante periodos de tiempo relativamente largos. Una de

sus ventajas es que permite evaluar el impacto de la intervención sobre varias enfermedades simultáneamente. Por ejemplo, también se evalúa el efecto de una dieta pobre en grasas sobre la incidencia de cáncer de colon (Ejemplo 5.2) y de la vacunación de la hepatitis B sobre la incidencia de otras enfermedades hepáticas crónicas (Ejemplo 5.3).

En los ensayos de campo, la unidad de asignación a la intervención puede ser un individuo (como en el Ejemplo 5.2) o, por otra parte, un grupo de personas. El grupo puede ser una familia, una escuela o toda una comunidad. Si la unidad de asignación es una comunidad (como en el Ejemplo 5.4), el estudio se denomina ensayo comunitario.

Ejemplo 5.4. *El Estudio de Intervención Comunitario de Deshabitación al tabaco [Community Intervention Trial for Smoking Cessation] es un estudio en curso diseñado para evaluar un programa de deshabitación al tabaco en los EE.UU. Se seleccionaron 11 pares de comunidades y se asignó al azar un miembro de cada par para recibir un programa de deshabitación al tabaco mientras el otro actuaba como control. La intervención se diseñó para favorecer la deshabitación del hábito tabáquico utilizando gran cantidad de recursos comunitarios para influir sobre las actitudes de la comunidad en relación al tabaquismo (COMMIT Research Group, 1991).*

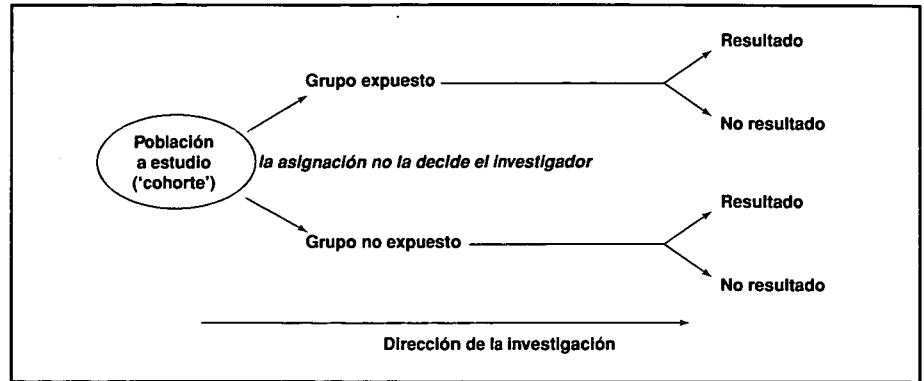
Los ensayos de intervención son muy potentes para probar hipótesis. A pesar de esto, no son el diseño de estudio más utilizado en epidemiología, principalmente debido a las limitaciones éticas.

5.1.2 Estudios de cohortes

Los estudios de cohortes son estudios observacionales en los que inicialmente se selecciona una población a estudio o cohorte. A continuación se obtiene información para determinar quién está expuesto y quién no al factor de interés. La cohorte se sigue a través del tiempo y se compara la incidencia de la enfermedad a estudio entre los individuos expuestos y los no expuestos (Figura 5.3). Este tipo de estudio observacional es el que más se parece a los estudios de intervención, a excepción de que la asignación de los sujetos a la exposición no está controlada por el investigador.

Ejemplo 5.5. *En 1980, 89.538 enfermeras de los EE.UU. entre 34 y 59 años y sin historia previa de cáncer, cumplimentaron un cuestionario dietético validado previamente y diseñado para medir el consumo individual de grasa total, grasa saturada, ácido linoleico y colesterol, así como otros nutrientes. Las enfermeras se clasificaron entonces en cinco grupos de igual tamaño según los niveles de su ingesta de grasa. Las mujeres de esta cohorte se siguieron en el tiempo y se comparó la incidencia de cáncer de mama entre los grupos (Willet et al., 1987).*

Figura 5.3.
Perfil de un estudio de cohorte.



Los estudios de cohortes seleccionan individuos y los clasifican según su estado de exposición. A veces pueden clasificarse simplemente como expuestos/no expuestos. Es más usual identificar varios grados de exposición. En el Ejemplo 5.5, las enfermeras (la 'cohorte') se clasificaron en cinco grupos (quintiles) según los niveles de ingesta de grasa, se siguieron en el tiempo y se comparó la incidencia de cáncer de mama ('resultado') entre los cinco grupos.

***Ejemplo 5.6.** Se inició un estudio de cohorte en varones, empleados por el gobierno chino en Taiwan, para investigar la asociación entre la infección por el virus de la hepatitis B y el carcinoma hepatocelular. Todos los participantes cumplieron un cuestionario y proporcionaron una muestra de sangre a la inclusión del estudio. Se incluyeron un total de 22.707 hombres: 3.454 eran positivos al antígeno de superficie de la hepatitis B (HBsAg) y 19.253 negativos. Estos hombres se siguieron y se comparó la incidencia de carcinoma hepatocelular entre los portadores de HBsAg y los no portadores (Beasley et al., 1981).*

En el Ejemplo 5.6, la cohorte consiste en hombres empleados por el gobierno chino en Taiwan. Los sujetos se clasificaron entre portadores del HBsAg ('expuestos') o no portadores ('no expuestos') en el momento de su inclusión en el estudio y se siguieron en el tiempo para evaluar si el riesgo de carcinoma hepatocelular ('resultado') era mayor en los expuestos en comparación con los no expuestos.

En los estudios de cohortes el investigador intenta asegurar que los grupos comparados sean lo más parecidos posible respecto a todos los factores que pueden estar relacionados con la enfermedad. Como el investigador no posee el control sobre quien está y quien no está expuesto, es probable que existan diferencias en relación a algunos factores (no investigados) entre los dos grupos, y en los análisis deberán utilizarse técnicas especiales para asegurar que se tengan en cuenta estas distribuciones desiguales (Capítulos 8, 13 y 14).

5.1.3 Estudios caso-control

Los estudios caso-control son estudios observacionales en los que el inicio es la identificación de 'casos' con la enfermedad (u otro evento de interés) de interés y de 'controles' adecuados sin esta enfermedad (u otro evento de interés). Entonces, los casos y controles se comparan para evaluar si existen diferencias en su anterior exposición a los factores de riesgo (Figura 5.4).

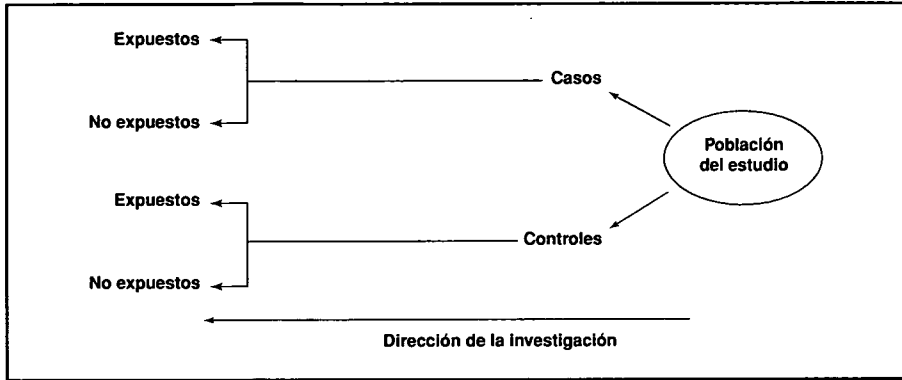


Figura 5.4. Perfil de un estudio caso-control.

Ejemplo 5.7. En Singapur se realizó un estudio caso-control para evaluar el papel de la dieta en el cáncer de mama. Doscientas mujeres chinas con cáncer de mama confirmado histológicamente y 420 controles sin esta enfermedad participaron en el estudio. Se utilizó un cuestionario de consumo de alimentos para estimar la ingesta dietética en el pasado. Se comparó la ingesta de alimentos y nutrientes entre los casos y controles, para evaluar si habían diferencias en el consumo (Lee et al., 1991).

En el Ejemplo 5.7, se identificaron las mujeres con ('casos') y sin ('controles') cáncer de mama y se comparó su dieta ('exposición') en el pasado.

Ejemplo 5.8. Se efectuó un estudio caso-control en Taiwan para evaluar si la infección por hepatitis B intervenía en la etiología del carcinoma hepatocelular. Un total de 128 casos de carcinoma hepatocelular confirmado por histología o citología y 384 controles sin la enfermedad se incluyeron en el estudio. Un 77% de los casos eran portadores del antígeno de superficie de la hepatitis B (HBsAg) en comparación a sólo un 28% de los controles (Chuang et al., 1992).

En el Ejemplo 5.8, se identificaron los sujetos con ('casos') y sin ('controles') carcinoma hepatocelular y se comparó la seropositividad del HBsAg ('exposición') en los dos grupos.

La principal diferencia entre los métodos de cohorte y caso-control está en la selección de los sujetos a estudio. Un estudio de cohorte selecciona sujetos inicialmente exentos de la enfermedad y los sigue en el tiempo. Por el contra-

rio, el método caso-control selecciona sujetos en función de la presencia o ausencia de la enfermedad (o cualquier otro estado patológico) en estudio.

Los estudios caso-control son especialmente adecuados para investigar enfermedades poco frecuentes tales como el cáncer. Un estudio de cohorte requiere el seguimiento de un gran número de individuos durante un largo periodo de tiempo para acumular casos suficientes de una enfermedad de este tipo. El método caso-control también es más apropiado para estudiar enfermedades con un largo periodo de inducción. Esto es debido a que los estudios caso-control se inician con sujetos que ya han desarrollado la enfermedad de interés, de modo que no es necesario esperar a que pase el tiempo entre una exposición y la manifestación de la enfermedad como en los estudios de cohortes.

El número de sujetos necesario para un estudio caso-control es mucho menor que el número requerido para los estudios de cohortes. Los Ejemplos 5.5 y 5.6 incluyen un gran número de personas expuestas durante periodos largos de tiempo. Las mismas cuestiones se tratan en los Ejemplos 5.7 y 5.8 utilizando un número de personas mucho menor. Así, los estudios caso-control son relativamente económicos, como mínimo en comparación con los estudios de cohortes.

En contrapartida, los resultados de los estudios caso-control son más difíciles de interpretar. En primer lugar, los controles deben obtenerse de la misma población a estudio de la que se obtuvieron los casos. En caso contrario, los resultados estarán distorsionados por sesgos de selección (véase Capítulo 9). En segundo lugar, no se puede estar seguro de que la exposición precediera a la enfermedad (excepto cuando la exposición es un atributo fijo, como el grupo sanguíneo, que no cambia con el tiempo). En el Ejemplo 5.7, es posible que la dieta descrita sea una consecuencia más que una causa del cáncer de mama.

5.1.4 Estudios transversales

En un estudio transversal, se selecciona una muestra de individuos de una población definida previamente y se recoge información sobre la(s) exposición(es) y la(s) enfermedad(es) de interés simultáneamente en un momento único del tiempo.

Ejemplo 5.9. Para evaluar si existe una asociación entre la infección por Helicobacter pylori y la gastritis atrófica crónica, condición relativamente común y precursora del cáncer de estómago, se efectuó un estudio transversal en 1.815 donantes de sangre sanos elegidos al azar en cuatro divisiones administrativas de Japón. Se tomaron muestras de sangre de todos los sujetos a estudio y se determinaron los anticuerpos IgG del Helicobacter pylori, y el pepsinógeno I y II (marcadores de gastritis crónica atrófica) en suero. La prevalencia de los anticuerpos en sujetos con gastritis atrófica crónica se comparó con la prevalencia en individuos sin gastritis atrófica crónica (Fukao et al., 1993).

En el Ejemplo 5.9 se seleccionó una muestra de 1.815 sujetos de una población bien definida (donantes de sangre sanos de divisiones administrativas de Japón determinadas) y se tomó una muestra de sangre en un momento concreto para determinar los anticuerpos del *H. pylori* ('exposición') y de pepsinógeno I y II ('resultado').

En este tipo de estudios es importante asegurarse de que la muestra de los sujetos que participan en el estudio sea representativa de la población total a la que se extrapolarán los resultados. De lo contrario, los resultados estarán distorsionados por un sesgo de selección. La mejor manera de evitar el sesgo de selección es utilizar métodos de muestreo aleatorios. Estos métodos aseguran que únicamente el azar determine quién se incluirá o no en la muestra (ver Capítulo 10).

Estos estudios son relativamente simples de realizar y requieren un corto periodo de tiempo debido a que no es necesario un seguimiento de los sujetos al estudio. Los estudios transversales se utilizan generalmente para estimar la prevalencia de enfermedades comunes de duración prolongada. Así, estos estudios no son apropiados para estudiar enfermedades tales como el cáncer, ya que sería necesario estudiar una población muy grande para identificar los casos suficientes y poder extraer conclusiones. Además, los casos prevalentes de cáncer son una muestra sesgada de todos los casos, ya que aquellos con supervivencias largas estarán sobrerrepresentados.

El uso principal de los estudios transversales en epidemiología del cáncer ha sido examinar la distribución y los determinantes de condiciones habituales, como la infección por el virus del papiloma humano o naevi cutáneo, que se conoce (o se sospecha) que están asociadas con el cáncer. Este tipo de estudios también se ha utilizado para investigar la distribución y los determinantes de comportamientos de alto riesgo conocidos (o potenciales), tales como ser fumador o ser un usuario habitual de aparatos de rayos UVA.

Estos estudios son relativamente simples de realizar, rápidos y baratos. Su principal inconveniente es que, al igual que en los estudios caso-control, no es posible saber si el evento siguió a la exposición en el tiempo o lo precedió, ya que la información sobre la exposición y el resultado se obtienen simultáneamente en el mismo momento. En el Ejemplo 5.9, no se puede establecer si la infección por *H. pylori* fue anterior o posterior a la gastritis atrófica crónica. Esto obviamente no es un problema para las exposiciones que no cambian con el tiempo tales como el sexo, raza o determinados rasgos genéticos. Para exposiciones que pueden variar, los estudios transversales pueden incluir cuestiones sobre el pasado así como exposiciones actuales. Por ejemplo, en un estudio sanitario de trabajadores de una industria determinada, puede preguntarse a los trabajadores sobre su trabajo actual y otros trabajos que hayan tenido en el pasado.

5.1.5 Estudios basados en datos recogidos de forma sistemática

El aspecto distintivo de este tipo de estudio es que los datos sobre la(s) exposición(es) y la(s) enfermedad(es) de interés se obtienen de sistemas de recopilación ya establecidos, sin que el investigador se ponga en contacto

con los sujetos. Los estudios basados en estos datos pueden efectuarse a nivel individual o agregado.

Nivel individual

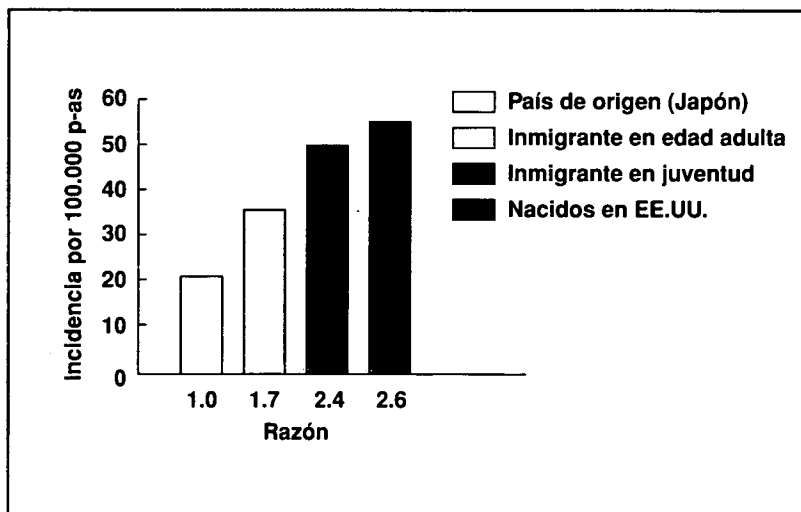
Muchos sistemas de recopilación de datos sistemáticos reúnen información sobre atributos personales tales como edad, sexo, lugar de nacimiento, lugar de residencia, ocupación, etc. La aparición de cáncer puede ser examinada según estas variables. El objetivo es buscar modelos que puedan indicar o confirmar hipótesis etiológicas específicas.

En el Ejemplo 5.10, los datos sobre el lugar de nacimiento, lugar de residencia y la edad de migración ('exposiciones'), y sobre el cáncer de mama ('resultado') estaban disponibles en los censos de población, archivos de la seguridad social y registros de cáncer para cada uno de los individuos del estudio.

El aspecto clave de este tipo de estudios es que los datos sobre la exposición y la enfermedad de interés se obtienen para cada uno de los sujetos de los sistemas de recopilación de datos existentes. No obstante, en cuanto a su análisis e interpretación pueden considerarse similares a los estudios de cohorte o caso-control. Por ejemplo, el Ejemplo 5.10 puede considerarse una cohorte de mujeres clasificadas en cuatro categorías diferentes de ex-

Ejemplo 5.10. El riesgo de cáncer de mama en mujeres se estudió en cuatro grupos de población: 1) mujeres japonesas nacidas y residentes en Japón; 2) mujeres japonesas que emigraron a los EE.UU. en su juventud; 3) mujeres japonesas que emigraron a los EE.UU. en edad adulta; y 4) mujeres nacidas en los EE.UU. Los principales resultados se muestran en la Figura 5.5. Éstos indican que los factores ambientales (p.e. dieta) en edades muy tempranas pueden ser importantes en el desarrollo de cáncer de mama (Shimizu et al., 1991).

Figura 5.5. Tasas de incidencia ajustadas por edad de mujeres con cáncer de mama residentes en los EE.UU. (1972-85) según el lugar de nacimiento y la edad de migración, y en Japón (1973-81) (reproducido con la autorización de Churchill Livingston, de Shimizu *et al.*, 1991).



posición según su lugar de nacimiento, residencia y edad de migración. Estas mujeres se siguieron en el tiempo y se determinó y comparó la aparición de cáncer de mama para cada grupo.

La principal ventaja de los estudios basados en estos datos respecto a otros estudios observacionales es que son relativamente rápidos y baratos ya que los datos están recopilados y no es necesario ponerse en contacto con los individuos. No obstante, su principal limitación es que las variables disponibles son limitadas y no se han obtenido específicamente para el estudio. En el Ejemplo 5.10, las variables 'país de nacimiento y residencia' son sólo medidas aproximadas de exposiciones más importantes desde el punto de vista biológico, tales como factores reproductivos, dieta y otras características del estilo de vida que no son recogidas sistemáticamente.

Nivel agregado (estudios ecológicos)

Los estudios en que se analiza la frecuencia de la enfermedad (o cualquier otro evento de interés) respecto al nivel de la exposición en varios grupos de individuos (o en el mismo grupo en diferentes periodos de tiempo) se denominan estudios ecológicos. En este tipo de estudios, no es posible relacionar la exposición de un individuo particular con el estado de salud. Por lo tanto, el grupo, más que el individuo, es la unidad de observación y análisis. Los grupos pueden definirse de muchas maneras según el lugar de residencia, lugar de nacimiento, situación socioeconómica, ocupación, etc.

La mayoría de los estudios ecológicos en epidemiología del cáncer utilizan datos obtenidos sistemáticamente. Los datos sobre la exposición media (o su frecuencia) en diferentes grupos de población pueden obtenerse de fuentes gubernamentales o privadas que recopilan datos sistemáticamente sobre variables demográficas, ambientales y de estilo de vida. Las tasas de enfermedad pueden obtenerse de programas de vigilancia, registros de cáncer o de certificados de defunción.

En el Ejemplo 5.11, no es posible relacionar la experiencia de cáncer de mama de cada mujer individual con su dieta debido a que la única información disponible para cada uno de los países fue la tasa de incidencia de cáncer de mama y una estimación de su consumo de grasa medio per cápita.

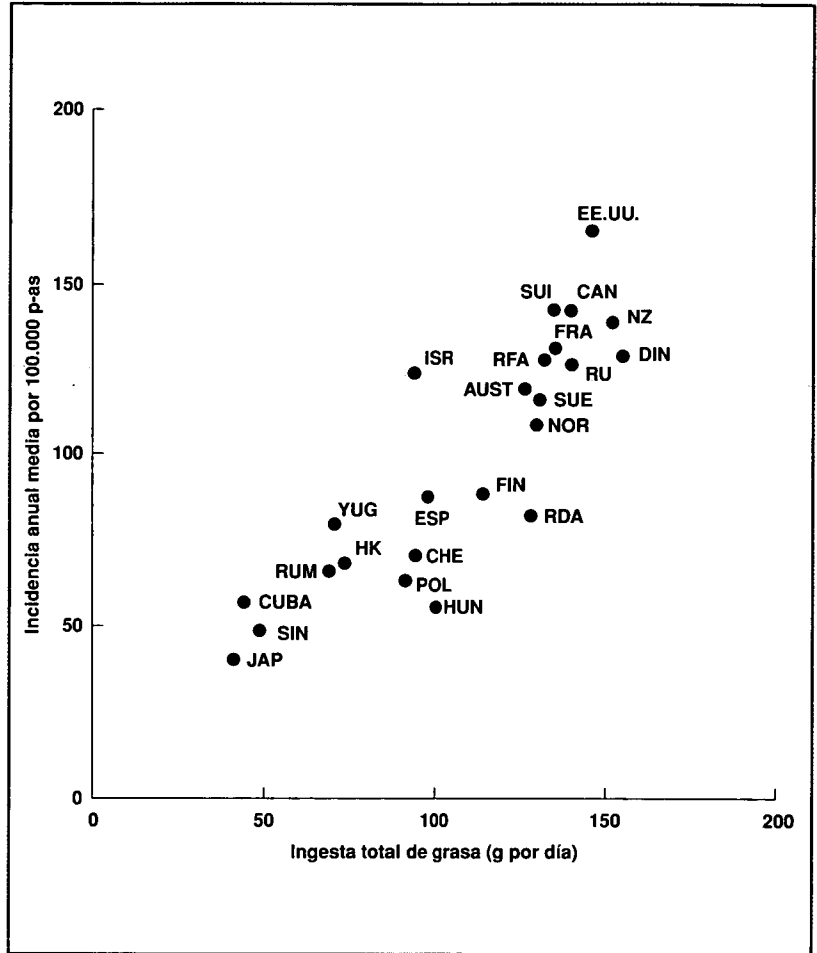
En la Figura 5.6, la incidencia de cáncer de mama (la 'enfermedad' ó el 'resultado') para diferentes países (los grupos en estudio) se examinó respecto a su consumo diario medio de grasa per cápita (la 'exposición').

A veces los datos sobre exposición(es) o resultado(s) de interés no están disponibles en los sistemas de recopilación de datos sistemáticos pero pueden obtenerse de estudios realizados anteriormente (como en el Ejemplo 5.12).

En el Ejemplo 5.12 se obtuvo información sobre la prevalencia de anticuerpos IgG del *H. pylori* ('exposición') y sobre la mortalidad por cáncer de estómago ('resultado') para cada una de las 46 regiones rurales (los grupos de interés) a partir de estudios realizados previamente en China.

Ejemplo 5.11. Los datos sobre la incidencia de cáncer de mama y el consumo diario medio per cápita de una amplia serie de alimentos de 24 países se obtuvieron de fuentes de datos recopilados sistemáticamente (Armstrong y Mann, 1985). La relación entre el consumo de grasas y el cáncer de mama se muestra en la Figura 5.6.

Figura 5.6. Consumo diario estimado de grasas per cápita en 1964-66 y tasas de incidencia de cáncer de mama en mujeres de 35-64 años en 1972-77 en 24 países (reproducido con la autorización de Oxford University Press, de Armstrong y Mann, 1985).



La principal ventaja de los estudios ecológicos es que son rápidos y baratos. Por ejemplo, la pregunta ¿aumenta una dieta rica en grasas el riesgo de cáncer de mama? puede tratarse de una manera más rápida y barata en un estudio ecológico que utilizando cualquier otro diseño (Ejemplos 5.2, 5.5 y 5.7).

Los estudios ecológicos son útiles para generar hipótesis. En contrapartida, tienen un limitado valor para evaluar la existencia de una relación real exposición-resultado a nivel individual, ya que sus resultados se refieren a grupos de personas y no pueden extrapolarse a individuos. Los intentos de inferir asociaciones a nivel individual a partir de las observadas en el grupo

Ejemplo 5.12. Se realizó un estudio para examinar la asociación entre la prevalencia de la infección por *Helicobacter pylori* y la mortalidad por cáncer de estómago en 46 regiones rurales de la República Popular de la China. Los resultados se representan gráficamente en la figura siguiente (Forman et al., 1990).

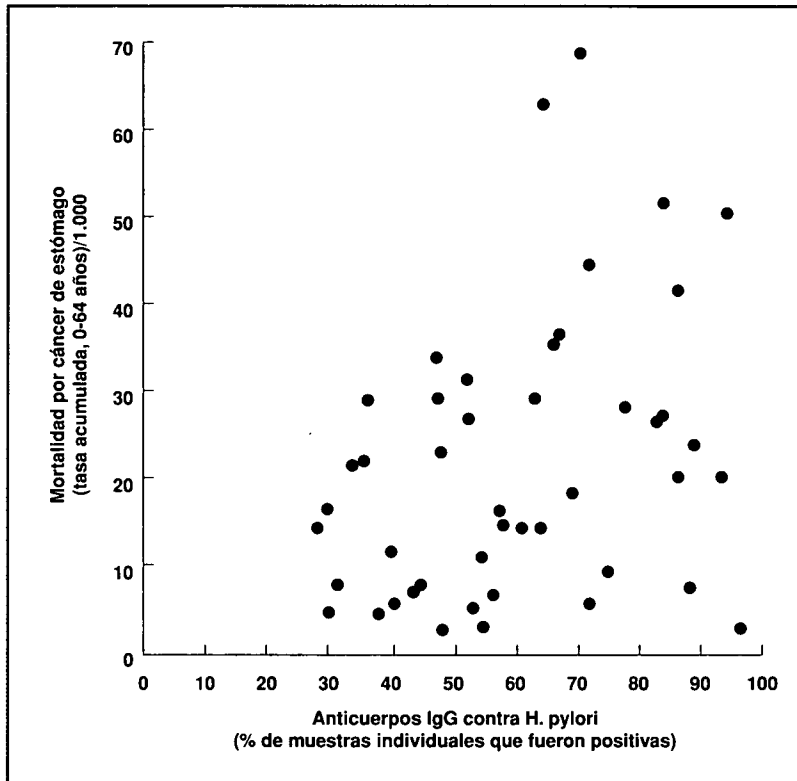


Figura 5.7.

Prevalencia de anticuerpos IgG del *H. pylori* y mortalidad por cáncer de estómago en 46 regiones rurales de la República Popular de China; cada región se representa por un punto en el gráfico (reproducido con la autorización de Wiley-Liss, Inc., subsidiaria de John Wiley & Sons, Inc., de Forman et al., 1990).

están sujetos a la 'falacia ecológica' (Apartado 11.2). Por ejemplo, en el Ejemplo 5.11 no podemos estar seguros de que las mujeres que desarrollaron cáncer de mama en cada país fueran realmente las que tomaban una dieta rica en grasas.

A pesar de estas limitaciones, los estudios ecológicos pueden ser la mejor estrategia para estudiar exposiciones que son más fáciles de medir en un grupo que individualmente, tales como la contaminación del aire y la calidad del agua. También son útiles para monitorizar la eficacia de las intervenciones en la población tales como las campañas de educación y los programas masivos de cribaje (Capítulo 16).

5.2 Medidas del efecto de la exposición

El objetivo básico de un estudio epidemiológico es cuantificar la asociación entre la exposición y la enfermedad o el resultado de interés. Esto requiere comparar la incidencia de la enfermedad en un grupo de individuos

expuestos al factor de riesgo estudiado con la incidencia de la misma enfermedad en un grupo de personas no expuestas. Esta comparación puede resumirse mediante la razón de las incidencias (u otras estimaciones de la frecuencia de la enfermedad para los dos grupos), que estima la probabilidad de desarrollar la enfermedad en los individuos expuestos respecto a los no expuestos, o mediante la diferencia entre las dos, que proporciona información sobre el efecto absoluto de la exposición en los expuestos comparado con los no expuestos.

5.2.1 Medidas relativas del efecto de la exposición

Estas medidas estiman la magnitud de la asociación entre la exposición y la enfermedad e indican cuán mayor es la probabilidad de que el grupo expuesto desarrolle la enfermedad respecto al no expuesto. Se pueden calcular tres tipos de medidas relativas:

$$\text{Razón de riesgos} = \frac{\text{Riesgo en el grupo expuesto}}{\text{Riesgo en el grupo no expuesto}}$$

$$\text{Razón de tasas} = \frac{\text{Tasa de incidencia en el grupo expuesto}}{\text{Tasa de incidencia en el grupo no expuesto}}$$

$$\text{Odds ratio (de la enfermedad)} = \frac{\text{Odds de la enfermedad en el grupo expuesto}}{\text{Odds de la enfermedad en el grupo no expuesto}}$$

Estas medidas muchas veces se denominan medidas de riesgo relativo.

El riesgo relativo se utiliza como una medida de fuerza etiológica. Un valor de 1,0 indica que las incidencias de la enfermedad en los grupos expuestos y no expuestos son idénticas y por tanto que no se observa una asociación entre la exposición y la enfermedad. Un valor mayor de 1,0 indica una asociación positiva, o un mayor riesgo entre los expuestos. De la misma forma, un riesgo menor de 1,0 muestra una asociación inversa o un menor riesgo entre los expuestos. En el Ejemplo 5.13, la probabilidad de que los trabajadores expuestos al riesgo ocupacional en cuestión murieran de cáncer era el doble que la de los no expuestos.

Obsérvese que en el ejemplo anterior las tres medidas proporcionan estimaciones similares del riesgo relativo. Esto es debido a que la muerte por cáncer es poco frecuente y por tanto el número de sujetos a riesgo permanece prácticamente constante a lo largo del estudio, ya que los casos representan una fracción despreciable de la población. En la práctica, las tres medidas del efecto proporcionarán estimaciones similares del riesgo relativo sólo para enfermedades de baja prevalencia (p.e. cáncer). Las estimaciones obtenidas con estas tres medidas pueden diferir considerablemente cuando se examina una enfermedad más frecuente (p.e. la mayoría de las

Ejemplo 5.13. Supongamos que 20.000 trabajadores se incluyeron en un estudio de cohorte. En el momento de su inclusión en el estudio, los individuos se clasificaron como expuestos o no expuestos a una sustancia en particular, basándose en el tipo de trabajo que tenían en ese momento. Toda la cohorte se siguió durante un periodo de dos años para establecer si los expuestos presentaban un mayor riesgo de morir de cáncer. Los resultados se presentan en la Tabla 5.1.

	Exposición	
	Sí	No
Nº inicialmente a riesgo	4.000	16.000
Muertes	30	60
Personas-años a riesgo ^a	7.970	31.940
Razón de riesgo = $\frac{30/4.000}{60/16.000} = \frac{7,5 \text{ por } 1.000}{3,75 \text{ por } 1.000} = 2,0000$		
Razón de tasas = $\frac{30/7.970 \text{ p-as}}{60/31.940 \text{ p-as}} = \frac{3,76 \text{ por } 1.000 \text{ p-as}}{1,88 \text{ por } 1.000 \text{ p-as}} = 2,0038$		
Odds ratio de la enfermedad = $\frac{30/(4.000-30)}{60/(16.000-60)} = \frac{0,00756}{0,00376} = 2,0076$		

^a Suponiendo que, en promedio, todas las muertes se produjeron en la mitad del periodo de seguimiento.

Tabla 5.1. Datos hipotéticos de una cohorte de 20.000 individuos seguidos durante un periodo de 2 años.

enfermedades infecciosas) o cuando se estudia una enfermedad moderadamente rara durante un largo periodo de tiempo (p.e. cardiopatía coronaria en mujeres seguidas durante 20 años).

5.2.2 Medidas absolutas del efecto de la exposición

La información sobre el riesgo relativo no proporciona una descripción completa de la asociación entre la exposición y la enfermedad. La Tabla 5.2 muestra los riesgos relativos, calculados como razón de tasas, de las enfermedades A y B entre los expuestos a un factor de riesgo determinado. Aunque la razón de tasas es mayor para la enfermedad A, la tasa de incidencia para la enfermedad A sólo aumenta en 15 casos por 100.000 personas-años a riesgo, mientras que la tasa de incidencia para la enfermedad B aumenta en 40 casos por 100.000 a riesgo. Está claro que el impacto absoluto de la exposición, medido por la diferencia de tasas, es bastante diferente para estas dos enfermedades.

El exceso de riesgo (también denominado riesgo atribuible) es una medida absoluta del efecto de la exposición. Indica cuántos casos pueden ser atribuibles a una exposición determinada, suponiendo que la relación en-

Tabla 5.2.
 Incidencia de las enfermedades A y B entre personas expuestas y personas no expuestas a un factor de riesgo determinado, razón de tasas y diferencias de tasas: datos hipotéticos.

	Enfermedad A	Enfermedad B
Tasa de incidencia en el grupo expuesto ^a	20	80
Tasa de incidencia en el grupo no expuesto ^a	5	40
Razón de tasas	4,0	2,0
Diferencia de tasas ^a	15	40

^a Tasas por 100.000 p-as.

tre la exposición y la enfermedad sea causal. Se puede calcular como la diferencia del riesgo (diferencia de riesgo) o de tasas (diferencia de tasas).

El exceso de riesgo es igual a:

Diferencia de riesgo = Riesgo en los expuestos – Riesgo en los no expuestos

ó

Diferencia de tasas = Tasa en los expuestos – Tasa en los no expuestos

El exceso de riesgo es especialmente útil al evaluar el impacto de la introducción o de la eliminación de un factor de riesgo y su valor puede expresarse como el número de casos de la enfermedad entre los expuestos que podrían prevenirse si la exposición se eliminara completamente. En el ejemplo anterior, la importancia de la exposición como factor etiológico estaba dada por la razón de tasas y fue mayor para la enfermedad A que para la enfermedad B, pero, desde un punto de vista de salud pública, la exposición es mucho más importante en la enfermedad B ya que se evitarían más casos de la enfermedad B que de la enfermedad A si se eliminara completamente la exposición (suponiendo que los costes en términos de salud pública de un caso de enfermedad A sean similares a los de la enfermedad B). Así, el riesgo relativo mide la fuerza de la asociación entre la exposición y la enfermedad de interés mientras que el exceso de riesgo mide el impacto de la asociación en términos de salud pública. A diferencia del riesgo relativo, la magnitud del exceso de riesgo no puede generalizarse a otras poblaciones porque depende de la incidencia basal en el grupo no expuesto, que tiende a variar de una población a otra.

Es útil expresar el exceso de riesgo con relación al riesgo (o tasa) en el grupo expuesto. Esta medida se denomina la fracción de exceso (también conocida como el *porcentaje de exceso de riesgo* o *porcentaje de riesgo atribuible*) y puede calcularse como:

$$\text{Fracción de exceso (\%)} = 100 \times (\text{exceso de riesgo / riesgo (o tasa) en el grupo expuesto})$$

En el ejemplo anterior, la fracción de exceso para las enfermedades A y B sería:

$$\text{Fracción de exceso (\%)} \text{ para la enfermedad A} = 100 \times (15 \text{ por } 100.000 \text{ p-as} / 20 \text{ por } 100.000 \text{ p-as}) = 75\%$$

Fracción de exceso (%) para la enfermedad B = $100 \times (40 \text{ por } 100.000 \text{ p-as} / 80 \text{ por } 100.000 \text{ p-as}) = 50\%$

Esta fracción de exceso representa la proporción de casos entre los expuestos que pueden atribuirse a la exposición (asumiendo causalidad). O para expresarlo de otra forma, representa la proporción de casos entre los expuestos que podrían haberse prevenido si nunca hubieran estado expuestos. Por tanto, un 75% de los casos de la enfermedad A y un 50% de los casos de la enfermedad B se habrían prevenido entre los expuestos si se hubiera eliminado el factor de riesgo en cuestión.

Ejemplo 5.14. Un total de 34.439 doctores británicos varones se siguieron durante 40 años y se evaluó su mortalidad en relación con el hábito de fumar (Doll et al., 1994a). La mortalidad por algunas enfermedades se muestra en la Tabla 5.3.

Causa de muerte subyacente	Nunca fumó regularmente Tasa ^b (1)	Fumador actual Tasa ^b (2)	Razón de tasas (2)/(1)	Diferencia de tasas ^b (2)-(1)	Fracción de exceso (%) (2)-(1) x 100 (2)
Cáncer					
Todos los tipos	305	656	2,2	351	54
Pulmón	14	209	14,9	195	93
Esófago	4	30	7,5	26	87
Vejiga	13	30	2,3	17	57
Enfermedades respiratorias (excepto cáncer)					
	107	313	2,9	206	66
Enfermedades vasculares					
	1.037	1.643	1,6	606	37
Todas las causas	1.706	3.038	1,8	1.332	44

^a Datos de Doll et al., 1994a.

^b Tasas ajustadas por edad por 100.000 p-as.

Tabla 5.3. Causas de mortalidad específica según el hábito de fumar en doctores británicos varones^a.

El Ejemplo 5.14 muestra que un 44% de las muertes que se produjeron entre los doctores británicos varones que fumaban podían atribuirse al tabaco (asumiendo la causalidad). La proporción de muertes que podían atribuirse al hecho de fumar variaron según la enfermedad. Para las enfermedades mostradas en la Tabla 5.3, la proporción más alta fue para el cáncer de pulmón (93%) y la más baja para las enfermedades vasculares (37%). No obstante, si los fumadores nunca hubieran fumado, el número total de muertes que se habrían evitado habría sido mucho mayor para las enfermedades vasculares (606 por 100.000 p-as) que para el cáncer de pulmón (195 por 100.000 p-as).

Se pueden calcular unas medidas similares cuando los expuestos presentan un menor riesgo de desarrollar la enfermedad que los no expuestos. En estas circunstancias, tendríamos:

Reducción de riesgo = riesgo (o tasa) en los no expuestos – riesgo (o tasa) en los expuestos

Fracción evitable = $100 \times (\text{reducción de riesgo} / \text{riesgo (o tasa) en los no expuestos})$

Ejemplo 5.15. Supongamos un grupo de mujeres que toma anticonceptivos orales y un grupo que nunca los ha tomado, se siguen en el tiempo y se mide y compara la incidencia de cáncer de ovario en ambos grupos. Los resultados de este estudio hipotético se muestran en la Tabla 5.4.

Tabla 5.4.

Distribución de los casos de cáncer de ovario y personas-años a riesgo en relación al uso de anticonceptivos orales: datos hipotéticos.

	Uso de anticonceptivos orales	
	Alguna vez	Nunca
Casos de cáncer de ovario	29	45
Personas-años a riesgo	345.000	321.429
Tasa por 100.000 p-as	8,4	14,0
Razón de tasas = 8,4 por 100.000 p-as/14,0 por 100.000 p-as = 0,60		
Reducción de riesgo = 14,0 – 8,4 por 100 000 p-as = 5,6 por 100.000 p-as.		
Fracción evitable (%) = $100 \times (5,6 \text{ por } 100.000 \text{ p-as} / 14,0 \text{ por } 100.000 \text{ p-as}) = 40\%$.		

En el Ejemplo 5.15, el 40% de los casos de cáncer de ovario podrían haberse prevenido si las mujeres que nunca consumieron anticonceptivos orales los hubieran usado.

5.3 Conclusiones

Antes de llevar a cabo un estudio, debe tenerse en cuenta cuál es el diseño más apropiado según la factibilidad práctica, la información a obtener, la duración esperada del estudio y los costes totales. Las ventajas e inconvenientes de cada uno de los diseños de estudio se tratarán con más detalle en otros Capítulos (7-11).

Todas las medidas de efecto comentadas en este capítulo pueden calcularse directamente en los estudios de intervención y de cohorte ya que se conoce la incidencia de la enfermedad en los expuestos y en los no expuestos. Esto no es así para otros diseños de estudios, para los cuales sólo pueden estimarse indirectamente. Por ejemplo, en los estudios caso-control los sujetos se seleccionan según si presentan la enfermedad ('casos') o no ('controles') y no según la presencia de exposición. Por tanto, no es posible calcular la incidencia en individuos expuestos y no expuestos. No obstante, es posible calcular la *odds* de la enfermedad entre casos y la *odds* de la exposición entre controles y obtener la *odds ratio* de la exposición (véase Capítulo 9). Puede demostrarse que, dependiendo del esquema de muestreo utilizado para seleccionar controles, esta medida proporciona una estimación no sesgada de una de las tres medidas del efecto consideradas en el Apartado 5.2.1. Esto se comentará con detalle en capítulos posteriores (9-11, 16).

Cuadro 5.1. Puntos esenciales

- Existen dos tipos principales de estudios epidemiológicos: de intervención (experimental) y observacional.
- En los *estudios de intervención*, la distribución de los sujetos a estudio en diferentes grupos la decide el investigador. Por tanto, si se realiza correctamente, los grupos serán similares en todos los aspectos a excepción de la exposición de interés. Existen dos tipos de estudio de intervención:
 - *Ensayos clínicos* donde el principal objetivo es evaluar el valor de nuevas formas de tratamiento.
 - *Ensayos de campo* donde el objetivo es evaluar si una intervención reduce el riesgo de la enfermedad entre las personas sanas. Los ensayos de campo pueden realizarse a nivel individual, si la unidad de asignación es el individuo, o a nivel agregado, si la unidad es un grupo de personas. Los ensayos comunitarios son un ejemplo de ensayos efectuados a nivel agregado en los que la unidad de asignación son comunidades enteras.
- En los *estudios observacionales*, los investigadores se limitan a observar la aparición de la enfermedad en personas que ya se han segregado en diferentes grupos en función de su exposición. Existen diferentes tipos de estudios observacionales:
 - *Estudios de cohortes* en los cuales se selecciona una población ('cohorte') y se determina la exposición de sus miembros. La cohorte se sigue en el tiempo y se mide y compara la aparición de la enfermedad en los diferentes grupos de exposición.
 - *Estudios caso-control* en los que se selecciona un grupo de pacientes con una enfermedad ('casos') y un grupo adecuado de sujetos sin esta enfermedad ('controles') y se determina y compara la exposición en el pasado a los factores de riesgo de interés.
 - *Estudios transversales* en los que se selecciona una muestra de sujetos de una población definida y se reúne información sobre la(s) exposición(es) y resultado(s) de interés simultáneamente en un único momento del tiempo.
 - *Estudios basados en datos sistemáticos* en los que los datos se obtienen de sistemas de recopilación de datos ya existentes (p.e. registro del cáncer o certificados de defunción). Pueden efectuarse a un nivel individual si se dispone de la información de la(s) exposición(es) y resultado(s) de interés para cada uno de los sujetos del estudio o a un nivel agregado (estudios ecológicos) si la unidad del estudio es el grupo en lugar del individuo.
- Una vez se han obtenido los datos de un estudio, la asociación entre la exposi-

Lecturas adicionales

* Las referencias a libros y artículos que tratan cada uno de los diseños de estudio se indicarán en cada uno de los capítulos correspondientes.

Cuadro 5.1. Puntos esenciales (Cont.)

ción y la enfermedad o resultado de interés puede cuantificarse por el cálculo de la *medida del efecto* apropiada. Ésta puede expresarse como la relación de la medida de la aparición de la enfermedad en los expuestos respecto a la misma en los no expuestos (medida relativa) o como la diferencia entre los dos (medida absoluta). El primer tipo de medida es especialmente importante al evaluar la etiología mientras que el segundo es más útil al evaluar el impacto sobre la salud pública.

- Cada diseño de estudio tiene sus propias limitaciones y ventajas. Éstas se considerarán con detalle en capítulos siguientes de este libro (Capítulos 7 a 13).

Capítulo 6

Evaluación del papel del azar

6.1 Poblaciones y muestras

Supongamos que como parte de una encuesta transversal de salud general, queremos determinar la proporción de hombres en una localidad particular que actualmente fuman cigarrillos. Debido a que por razones prácticas y económicas es imposible entrevistar a cada hombre en la localidad, decidimos seleccionar una muestra aleatoria de 30 hombres. En esta muestra la proporción de fumadores actuales es $7/30 = 23\%$.

Normalmente es imposible examinar cada individuo de una población de interés y estamos limitados a examinar una *muestra (o subgrupo) de individuos* tomada de esta población con la confianza de que sea representativa de la población.^a

Sin información de la población entera, no podemos saber la proporción real en la población. No obstante, la proporción calculada a partir de una muestra aleatoria puede utilizarse como un estimador de la proporción en toda la población de la que se tomó la muestra. En el ejemplo anterior, la proporción muestral (23%) es la mejor estimación de la proporción real, pero desconocida, de los fumadores actuales en toda la localidad.

6.1.1 ¿En qué medida es buena nuestra proporción muestral?

La muestra aleatoria concreta que hemos utilizado no tiene ninguna propiedad especial, y muestras aleatorias diferentes proporcionarán unas estimaciones ligeramente diferentes de la proporción de fumadores en la población. Esto implica que nuestro estimador muestral está sujeto a un *error de muestreo*. La proporción de los fumadores actuales en toda la localidad es poco probable que sea exactamente el 23% como se ha encontrado en nuestra muestra debido a la presencia del error de muestreo. La pregunta es, ¿a qué distancia del 23% es probable que esté?

Para intentar responder a la pregunta anterior, primero aceptamos que la muestra elegida sólo era una de las múltiples muestras posibles de tamaño 30. Supongamos que podemos estudiar 100.000 muestras. Para cada muestra, entrevistamos 30 individuos y calculamos la proporción de la muestra de fumadores actuales p . El valor de p variará de muestra a muestra. Si representamos gráficamente los valores de p , obtendremos una distribución como la mostrada en la **Figura 6.1**.

Esta distribución se denomina *distribución muestral de p* . Indica que, aunque es probable que la mayoría de las estimaciones muestrales se concentren alrededor de la proporción real (poblacional) π , algunas estarán lejos de este valor real. La amplitud de la dispersión nos indica cuán precisa

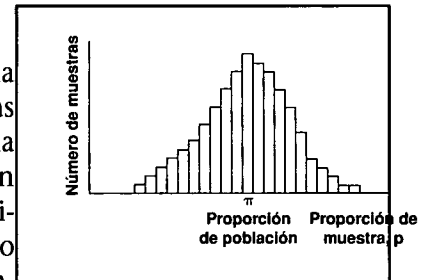


Figura 6.1. Distribución muestral de p para 100.000 muestras repetidas de tamaño 30.

^a Para asegurar la representatividad la muestra de individuos se debe seleccionar aleatoriamente de la población de interés. Es decir, cada individuo en la población debe tener la misma probabilidad de ser incluido en la muestra. Las diferentes maneras de tomar una muestra aleatoria de una población específica se tratarán en el Capítulo 10.

es nuestra proporción muestral p , como una estimación de la proporción real π . Si la distribución es amplia, existe mucho error de muestreo y nuestra p puede variar mucho respecto al valor real π . Si es estrecha, existe poco error de muestreo, y nuestra p es probable que esté muy próxima a π .

Ya hemos visto en el Apartado 3.3.1 que la dispersión de una distribución puede medirse por una cantidad denominada desviación estándar. Puede demostrarse que la desviación estándar de una distribución muestral de p viene dada por

$$EE(p) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

donde n representa el tamaño de la muestra. EE es el *error estándar*, que es el término que utilizamos generalmente para la desviación estándar de una distribución muestral. El error estándar nos indica la precisión con que nuestro valor p de la muestra estima el valor real, π . Observe que si aumentamos el tamaño de la muestra, n , disminuimos el error estándar y la distribución muestral será estrecha. Esto es justo lo que sería de esperar, ya que muestras más grandes deberían proporcionar estimaciones de π más precisas.

Cuando realizamos la encuesta, naturalmente, no conocemos el valor de π (¡de lo contrario no necesitaríamos hacer la encuesta!), y por tanto no podemos utilizar realmente la fórmula anterior para calcular el EE (p). Podemos obtener una buena estimación sustituyendo π en la fórmula con nuestro estimador muestral p , dando

$$EE(p) = \sqrt{\frac{p(1-p)}{n}}$$

Que puede reordenarse como

$$EE(p) = \sqrt{\frac{p^2(1-p)}{a}}$$

donde a es el numerador de la proporción $p = a/n$ (en nuestra muestra $a = 7$, el número observado de fumadores actuales). Esta última fórmula es especialmente útil porque muestra que el error estándar es inversamente proporcional al número observado de casos. Es el número de casos en el numerador de p el que determina básicamente la magnitud del error estándar y no el tamaño de la muestra de por sí.

Puede demostrarse matemáticamente que, en muestras de tamaño lo suficientemente grande, aproximadamente un 95% de todos los estimadores muestrales se encontrarán dentro de $\pm 1,96$ errores estándar alrededor del valor real π ; un 2,5% de los estimadores muestrales (uno de cada 40) será inferior a 1,96 EE por debajo del valor real y un 2,5% (uno de cada 40) será superior a 1,96 EE por encima del valor real (Figura 6.2).

¿Qué podemos decir ahora de π a partir de una muestra única de 30 individuos? Nuestra muestra puede proceder de cualquier parte de la distribución mostrada en la Figura 6.2. Ahora bien, antes de tomar la muestra había una probabilidad del 95% de que la proporción muestral p observada se encontrara dentro de un rango de más/menos dos errores estándar (con más precisión $1,96 \times EE$) del valor real π . Como consecuencia lógica, los intervalos calculados a partir de muestras de tamaño similar pero centrados en cada proporción muestral p incluirán π si p se encuentra entre dos errores estándar de π . Por tanto, un intervalo acotado entre los siguientes límites superior e inferior

$$p - 1,96 \times EE \text{ y } p + 1,96 \times EE$$

normalmente escrito $p \pm 1,96 \times EE$ (p) incluirá la proporción real π con una probabilidad del 95%. Estos límites, calculados a partir de los datos de la muestra, se denominan *límites de confianza inferior y superior*, respectivamente, y el intervalo entre ellos se denomina *intervalo de confianza al 95%* de la proporción poblacional desconocida π .

En nuestro ejemplo para la estimación de la proporción de hombres fumadores actuales,

$$n = 30, a = 7 \text{ y } p = 0,23$$

Estimamos que el error estándar p es

$$EE(p) = \sqrt{\frac{0,23^2(1-0,23)}{7}} = 0,076$$

Por tanto, un intervalo de confianza al 95% para la proporción real de hombres que fuman actualmente en toda la localidad viene dado por

$$0,23 \pm 1,96 \times 0,076 = 0,081 \text{ a } 0,379$$

Así nuestra mejor estimación de la proporción de fumadores actuales en toda la localidad es del 23%, pero el valor real podría estar fácilmente entre el 8% y el 38%.

En términos estrictos, el intervalo de confianza es un rango de valores que probablemente contenga el valor real de la población pero todavía no estamos seguros de que así sea. En realidad, un intervalo de confianza de

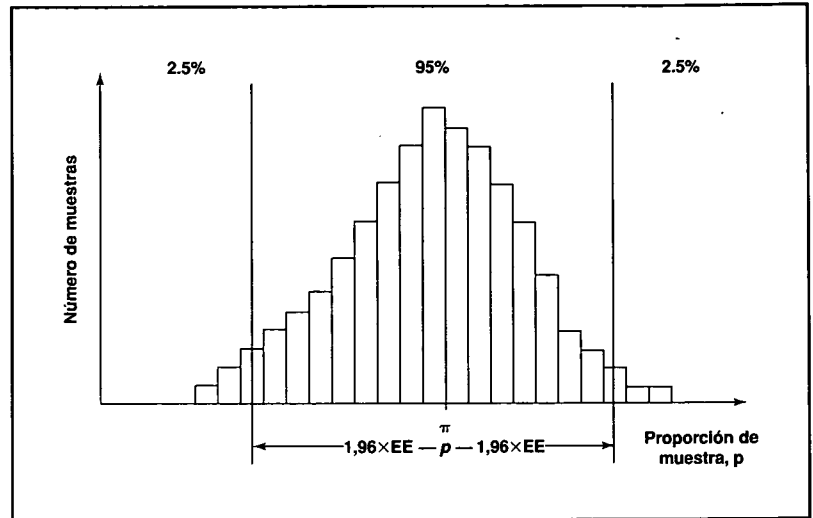


Figura 6.2. Distribución muestral de p con límites de confianza al 95%.

un estudio particular puede incluir o no el valor real de la población. El intervalo de confianza se basa en el concepto de repetición del estudio considerado. Por tanto, si el estudio se repitiera 100 veces, esperaríamos que 95 de los 100 intervalos de confianza al 95% resultantes incluyeran el valor real de la población.

Si queremos estar todavía más seguros de que nuestro intervalo incluye el valor real de la población, podemos calcular un *intervalo de confianza al 99%*. Esto se efectúa simplemente por sustitución del 1,96 por 2,58 en la fórmula anterior. Esto es, utilizamos

$$p \pm 2,58 \times \sqrt{\frac{p^2(1-p)}{a}}$$

En nuestro ejemplo, el intervalo de confianza al 99% correspondiente para la proporción de fumadores actuales en toda la localidad es 0,034 a 0,426, o del 3% al 43%.

Asimismo, podemos calcular un intervalo de confianza al 90% por sustitución de 1,96 por 1,64 en la fórmula

$$p \pm 1,64 \times \sqrt{\frac{p^2(1-p)}{a}}$$

En nuestro ejemplo, el intervalo de confianza al 90% correspondiente para la proporción de fumadores actuales en toda la localidad es 0,105 a 0,355, o del 11% al 36%.

6.1.2 ¿Cómo son de buenos los estimadores muestrales?

Esta manera de describir el error de muestreo es muy útil y no está limitada a la proporción muestral. Podemos obtener intervalos de confianza para cualquier otro estimador muestral como medias, riesgos, tasas, razones de tasas, diferencias de tasas, etc. El concepto fundamental es similar al mostrado anteriormente para las proporciones. En todos estos casos, los intervalos de confianza indican la precisión, en términos de probabilidad, de que nuestro estimador muestral sea el valor real de la población.

Supongamos que, como parte de la misma encuesta de salud general, los investigadores quieren determinar la altura media de los hombres en la misma localidad. Midieron los 30 individuos de su muestra y obtuvieron una altura media muestral de 165 cm. Otra vez, es poco probable que la altura media de la población masculina adulta en toda la localidad sea exactamente 165 cm. No obstante, si se tomara un número muy grande de muestras de 30 individuos de esta población y se calculara y representara gráficamente para cada uno la altura media, se obtendría una distribución muestral de la media. Esta distribución muestral tendrá una forma similar a la de la distribución de muestreo de una proporción (Figura 6.1), es decir, la distribución tendrá 'forma de campana' con la mayoría de las estimaciones muestrales centradas alrededor de la media de la población. Por tanto, po-

demos obtener el intervalo de confianza al 95% del mismo modo que el utilizado para las proporciones:

$$\text{Media de la muestra} \pm 1,96 \times \text{EE de la media}^b$$

El *error estándar de una media* puede estimarse mediante la fórmula:

$$\text{EE (media)} = \text{DE} / \sqrt{n}$$

donde *DE* representa la desviación estándar descrita en el Apartado 3.3.1. Supongamos que, en el ejemplo anterior, la desviación estándar obtenida era 7,1 cm. El error estándar de la media vendrá dado por

$$\text{EE (media)} = 7,1 / \sqrt{30} = 1,3 \text{ cm}$$

El intervalo de confianza al 95% para esta muestra será igual a

$$165 \text{ cm} \pm 1,96 \times 1,3 \text{ cm} = 162 \text{ cm a } 168 \text{ cm}$$

¿Cómo interpretamos este intervalo de confianza? Si el estudio se repitiera 100 veces, de los 100 intervalos de confianza al 95% resultantes, esperaríamos que 95 incluyeran el valor real de la media de la altura de la población. Por tanto, el intervalo de confianza de esta muestra en particular de 30 hombres proporciona un rango de valores que es probable que incluya la media real de la población aunque no podemos asegurar que lo haga.

Mientras la distribución muestral de una estimación particular tenga aproximadamente 'forma de campana' (es decir, lo que los estadísticos denominan 'Distribución normal'), y siempre lo será si el tamaño de la muestra es lo suficientemente grande, siempre podemos resumir el cálculo de un intervalo de confianza al 95% como sigue:

$$\text{estimador muestral} \pm 1,96 \times \text{EE (estimador muestral)}$$

(Para obtener un intervalo de confianza al 90% o al 99%, todo lo que hay que hacer es sustituir el 1,96 en la fórmula por 1,64 o 2,58, respectivamente).

En el Ejemplo 6.3e estimó que los hombres empleados durante 10 o más años tenían un exceso de 92 muertes por cáncer por 10.000 personas-años en comparación con los empleados durante menos de un año, con un intervalo de confianza al 95% que oscila de 61 a 122 muertes por 10.000 personas-años. Este intervalo de confianza se calculó utilizando la fórmula general anterior como sigue

$$\text{diferencia de tasas} \pm 1,96 \times \text{EE (diferencia de tasas)},$$

donde el EE de la diferencia de tasas era aproximadamente 15 muertes por 10.000 personas-años.

^b El valor preciso a utilizar en esta fórmula varía con el tamaño de la muestra y se indica en tablas de la distribución *t*. No obstante, para tamaños de muestra grandes (≥ 30) este factor es próximo a 1,96.

Ejemplo 6.1. Consideremos un ejemplo hipotético de un estudio de cohorte de 15.326 hombres empleados en una fábrica determinada. La tabla siguiente muestra su mortalidad por cáncer en función de la duración de su empleo (Tabla 6.1).

Tabla 6.1. Razones de tasas y diferencias de tasas ajustadas por edad de cáncer (todos los tipos combinados) en relación a la duración del empleo: datos hipotéticos.

Duración del empleo (años)	N.º de casos	Personas -años	Tasa ^a	Razón (IC 95%) ^b	Diferencia de tasas ^a (IC 95%) ^b
<1 ^c	44	40.056	11	1,0	0
1,0-1,9	67	21.165	32	2,9 (1,9-4,3)	21 (12-29)
2,0-4,9	19	3.105	61	5,6 (3,1-9,7)	50 (23-78)
5,0-9,9	48	5.067	95	8,6 (5,6-13,3)	84 (57-111)
≥10	43	4.192	103	9,3 (6,0-14,6)	92 (61-122)

^a Tasas por 10.000 personas-años.

^b IC = intervalo de confianza.

^c Categoría de referencia.

La razón de tasas correspondiente fue 9,3 con un intervalo de confianza al 95% que oscila de 6,0 a 14,6. Así, nuestra mejor estimación es que los hombres empleados durante 10 ó más años tenían una probabilidad nueve veces mayor de morir de cáncer que los empleados durante menos de un año, pero la razón de tasas real puede encontrarse entre 6,0 y 14,6 (y es poco probable que se encuentre fuera de este intervalo).

Puede haber observado en este ejemplo que mientras los límites de confianza de una diferencia de tasas son equidistantes del estimador muestral, esto no es así para los límites de confianza de una razón de tasas. Esto puede observarse claramente en las Figuras 6.3 y 6.4a. Al contrario de la diferencia de tasas, la distribución muestral de una razón de tasas no es simétrica ya que el menor valor posible que puede tomar es cero, mientras que el máximo es el infinito. Para obtener una distribución más simétrica, se utilizó una transformación logarítmica de los datos. Como consecuencia de esta transformación, los límites de confianza son equidistantes respecto al estimador muestral de la razón de tasas en la escala logarítmica (Figura 6.4b) pero asimétrica cuando se vuelve a convertir a la escala original (Figura 6.4a) (véase Anexo 6.1, al final de este capítulo, que proporciona fórmulas para calcular intervalos de confianza para las diferencias y razones de tasas).

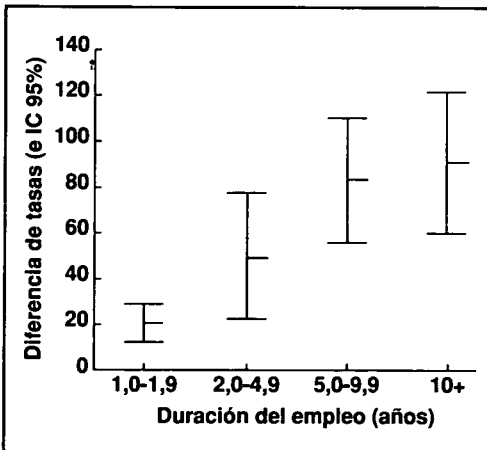


Figura 6.3. Representación gráfica de las diferencias de tasas (indicadas por las líneas horizontales medias) y sus intervalos de confianza al 95% (líneas verticales) en una escala aritmética (datos de la Tabla 6.1).

6.1.3 Representación de los intervalos de confianza

Si tenemos dos o más grupos, podemos representar las estimaciones muestrales y sus intervalos de confianza al 95% en un gráfico. Por ejemplo, las razones de tasas y las diferencias de tasas de la Tabla 6.1 y sus respectivos intervalos de confianza se representan en las Figuras 6.3 y 6.4.

Las líneas horizontales centrales muestran las diferencias de tasas observadas y las razones de tasas observadas, mientras que las líneas verticales indican los intervalos de confianza al 95%. Observe como el intervalo de confianza es mucho más estrecho cuanto mayor es el número de casos (p.e. categoría 1 – 1,9 años (basado en 67 casos)). Es el número de casos en el numerador de las tasas y las proporciones el que determina el tamaño del error estándar y, por tanto, la amplitud del intervalo de confianza.

6.1.4 Otros comentarios

La inferencia estadística es un proceso por el cual se extraen conclusiones sobre la población a partir de resultados observados en una muestra. Los métodos estadísticos descritos suponen que la muestra de individuos estudiada se ha seleccionado aleatoriamente a partir de alguna población de interés, que se definió previamente. Esto es, cada individuo en la población tiene la misma probabilidad de estar en la muestra elegida. Con bastante frecuencia en epidemiología, es imposible obtener una muestra realmente aleatoria y por tanto hay que considerar la posibilidad de *sesgo de selección* (Capítulo 13). Los intervalos de confianza indican sólo los efectos de la variación de muestreo sobre la precisión de las estimaciones de la muestra y no pueden controlar los errores que no son de muestreo tales como el sesgo en la selección de los sujetos del estudio o en la medida de las variables de interés. Por ejemplo, si la encuesta sobre el hábito de fumar sólo incluyera hombres que visitaron a sus médicos por problemas respiratorios, la muestra no sería representativa de toda la población masculina en la comunidad. Las técnicas estadísticas descritas anteriormente suponen que este sesgo no está presente.

Otro aspecto a considerar en estudios epidemiológicos es que, incluso cuando se estudian poblaciones enteras, la variabilidad aleatoria todavía debe ser considerada. Las tasas de mortalidad pueden calcularse a partir de estadísticas vitales nacionales, o las tasas de incidencia se pueden determinar a partir de registros de cáncer que incluyen poblaciones enteras, pero incluso en este caso habrá una variación aleatoria en el número de casos de un año a otro.

Ejemplo 6.2. La Tabla 6.2 muestra que existe una fluctuación aleatoria considerable en el número de casos de cáncer de labio en mujeres registradas de año en año en Inglaterra y Gales.

Año de registro										
1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
58	57	60	46	57	64	53	47	51	58	62

^a Datos de OPCS (1983a a 1994b).

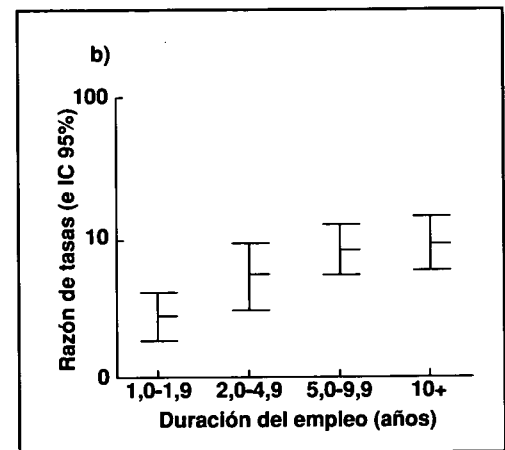
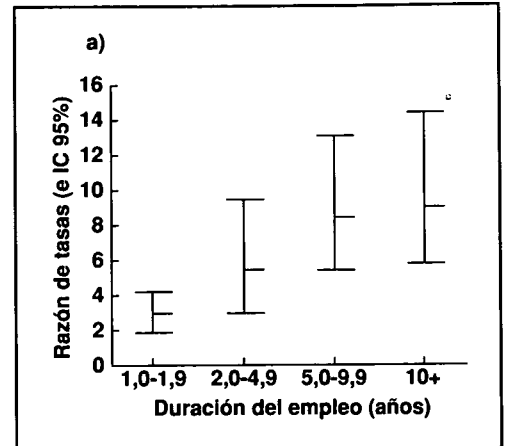


Figura 6.4. Representación gráfica de las razones de tasas y sus intervalos de confianza al 95%: (a) en escala aritmética y (b) en escala logarítmica (datos de la Tabla 6.1).

Tabla 6.2. Número de casos incidentes de cáncer de labio por año de registro, mujeres, Inglaterra y Gales, 1979-1989^a.

Aunque para producir los datos en el Ejemplo 6.2 se estudió toda la población del país, todavía había variabilidad aleatoria en el número de casos de cáncer de labio de año en año que no puede explicarse en términos de cambios del riesgo subyacente. En estas situaciones se está haciendo un muestreo 'en el tiempo', es decir, la población en un año cualquiera puede considerarse como una 'muestra'. Los métodos comentados anteriormente deben seguir utilizándose para evaluar el grado de precisión de las tasas observadas. Por supuesto, es posible que no toda la variación de año en año sea aleatoria y, por ejemplo, puede haber una tendencia, ascendente o descendente, durante un periodo de tiempo concreto (véase Apartado 4.3.2).

Cuando las tasas se basan en pocos casos, de modo que los intervalos de confianza son muy amplios, puede ser útil reducir la variación aleatoria reuniendo los datos de cinco, o incluso diez años consecutivos y calcular el promedio anual de las tasas de incidencia (o mortalidad).

6.2 Probando hipótesis estadísticas

Un investigador que recopila datos generalmente tiene dos objetivos: (a) extraer conclusiones sobre el valor real en la población, basándose en la información proporcionada por la muestra y (b) en segundo lugar probar hipótesis sobre el valor en la población.

Ejemplo 6.3. Supongamos que en un ensayo clínico sobre cáncer de cuello uterino con metástasis, 218 pacientes se distribuyeron aleatoriamente a un nuevo régimen de tratamiento o al tratamiento estándar. Todas las pacientes se siguieron durante un año tras su inclusión en el ensayo (o hasta la muerte si se producía antes). El número de mujeres todavía vivas al final de este periodo de seguimiento se presenta en la Tabla 6.3.

		Tipo de tratamiento	
		Nuevo	Estándar
Vivas al final del primer año	Sí	68	45
	No	40	65
Total		108	110

Tabla 6.3.

Número de pacientes todavía vivas un año después de su inclusión en el ensayo en función del tipo de tratamiento administrado: datos hipotéticos.

Los datos del Ejemplo 6.3 muestran que un 63,0% (68 de 108) de las pacientes a las que se había administrado el nuevo tratamiento todavía estaban vivas en comparación con un 40,9% (45 de 110) de las que recibieron el tratamiento estándar. A partir de estos resultados el nuevo tratamiento parece superior, pero ¿qué fuerza tiene la evidencia de que esto sea así?

En general, al comparar los efectos de dos niveles de una exposición o de dos tratamientos, se forman dos grupos, los 'expuestos' y los 'no expuestos' y se calculan sus estadísticas de resumen respectivas. Podemos querer comparar las dos muestras y preguntar: '¿Podrían las dos proceder de la misma población?' Es decir, ¿influye en el resultado que estamos observando el

hecho de que algunos sujetos estuvieran expuestos y otros no? Si no existe una evidencia fuerte a partir de datos de que la exposición influya en el desenlace, podemos suponer que las dos muestras proceden de la misma población respecto al desenlace particular considerado.

Una prueba de significación estadística es un método que intenta cuantificar la probabilidad de obtener los resultados observados si *no existe una diferencia real entre los grupos* examinados en la población general.

6.2.1 La hipótesis nula

Un investigador normalmente tiene algunas hipótesis generales (o teóricas) sobre el fenómeno de interés antes de iniciar un estudio. Por esto, en el

Ejemplo 6.3 se cree que el nuevo tratamiento puede ser mejor que el estándar en el tratamiento de pacientes con cáncer de cuello uterino con metástasis. Esta hipótesis se conoce como la *hipótesis del estudio*. No obstante, es imposible demostrar la mayoría de las hipótesis de manera concluyente. Por ejemplo, se puede formular la hipótesis de que todos los niños chinos tienen el pelo negro. Lamentablemente, incluso una vez observado un millón de niños chinos y comprobado que todos tienen pelo negro, ello no demostraría la hipótesis. Por otro lado, con sólo que se vea un niño chino de pelo rubio, la teoría se rechaza. Así, muchas veces hay una situación lógica más simple para rechazar hipótesis que para aceptarlas. Obviamente, la situación es mucho más compleja en epidemiología. La observación de que los dos tratamientos, el nuevo y el estándar, tienen efectos similares en un único paciente no sería suficiente para rechazar la hipótesis del estudio de que los dos tratamientos tenían un efecto diferente.

Volvamos a considerar el ensayo clínico de cáncer de cuello uterino (Ejemplo 6.3). Además de la hipótesis del estudio que propone que el nuevo tratamiento es mejor que el estándar, consideramos la *hipótesis nula* que propone que los dos tratamientos son igualmente eficaces. Si la hipótesis nula fuera cierta, entonces para la población de todas las pacientes con cáncer de cuello uterino con metástasis, la experiencia de supervivencia en un año sería similar para los dos grupos de pacientes, independientemente del tipo de tratamiento que recibieran. La formulación de una hipótesis nula, es decir, un planteamiento de que no existe una relación estadística verdadera, es el primer paso en cualquier prueba de significación estadística.

6.2.2 Prueba de significación

Tras especificar la hipótesis nula, la principal cuestión es:

Si la hipótesis nula fuera cierta, ¿cuál es la probabilidad de obtener una diferencia como mínimo tan grande como la observada?

Por ejemplo, en el ensayo del cáncer de cuello uterino, ¿cuál es la probabilidad de obtener una diferencia entre tratamientos como mínimo tan grande como el $63,0\% - 40,9\% = 22,1\%$ observado? Esta probabilidad, habitualmente indicada como P (P mayúscula mejor que la p minúscula utili-

zada anteriormente para la proporción de la muestra), se determina aplicando una prueba estadística apropiada.

Un ejemplo simple

Para entender la base para una prueba de significación estadística, veamos primero un ejemplo que utiliza números lo suficientemente pequeños como para permitir cálculos fáciles. Supongamos que un investigador tiene una teoría que predice que habrá un exceso de varones entre niños resultantes de una fertilización *in vitro* (FIV) y, por tanto, quiere estudiar la pregunta «¿Hay más niños que niñas entre los nacidos de FIV?».

El investigador formula una hipótesis nula de que existe una proporción igual (0,5 ó 50%) de niños y niñas en la población de nacidos de FIV. A continuación, toma una muestra de cinco historias clínicas de una clínica de FIV y observa que todos son niños en esta muestra de nacimientos. Ahora podemos calcular la probabilidad de obtener cinco niños y ninguna niña si la hipótesis nula de número igual de niños y niñas fuera cierta.

Probabilidad de que la primera muestra tomada sea un niño	= 0,5	
Probabilidad de que la primera y segunda muestras sean niños	= 0,5 x 0,5	= 0,25
Probabilidad de que la primera, segunda y tercera muestras sean niños	= 0,5 x 0,5 x 0,5	= 0,125
Probabilidad de que la primera, segunda, tercera y cuarta muestras sean niños	= 0,5 x 0,5 x 0,5 x 0,5	= 0,0625
Probabilidad de que las cinco muestras sean niños	= 0,5 x 0,5 x 0,5 x 0,5 x 0,5	= 0,03125

Por tanto, existe una probabilidad del 3,125% de obtener cinco niños seguidos, incluso aunque las proporciones reales de niños y niñas nacidos en toda la población eran iguales. ¡Acabamos de realizar una prueba de significación estadística! Este resultado indica la probabilidad (*P*) de obtener un resultado tan grande o mayor que el observado si no existieran diferencias reales entre la proporción de niños y niñas en toda la población de nacidos de FIV.

¿Qué podemos concluir de esta probabilidad? Este valor de *P* puede considerarse como una medida de la consistencia del resultado observado con la hipótesis nula. Cuanto menor sea el valor de *P*, más fuerte será la evidencia proporcionada por los datos contra la hipótesis nula. En este ejemplo, la probabilidad de obtener cinco niños seguidos, si la hipótesis nula fuera realmente cierta, era bastante baja. Por tanto, nuestros datos proporcionan una evidencia moderada de que el número de niños es mayor que el número de niñas entre los nacidos de FIV.

No obstante, a pesar de esta baja probabilidad, la hipótesis nula puede ser cierta. Nuestra interpretación final y conclusiones dependen mucho de nuestros conocimientos previos del fenómeno que estamos examinando. (La situación puede compararse con tirar una moneda. Incluso tras sacar cinco cruces en una serie de cinco tiradas la mayoría de la gente seguiría creyendo en la hipótesis nula de una moneda no sesgada. No obstante, si las 20 primeras tiradas fueran todas cruces los investigadores sospecharían

mucho de un sesgo ya que la probabilidad de que pase esto al azar es sólo de $(0,5)^{20} = 0,00000095$).

Comparando dos proporciones

En el estudio clínico de cáncer de cuello uterino (Ejemplo 6.3), si la hipótesis nula de que no existen diferencias en la supervivencia entre los dos tratamientos es cierta, ¿cuál es la probabilidad de encontrar una diferencia en la muestra tan grande o mayor que 22,1% (= 63,0% – 40,9%)? Si la hipótesis nula fuera cierta, la única razón para que la diferencia observada fuera mayor de cero es el error de muestreo. En otras palabras, incluso aunque la diferencia real en las proporciones de la población fuera cero, no esperaríamos que la diferencia de nuestra muestra en particular fuera exactamente cero debido al error de muestreo. En estas circunstancias, ¿cuánto, en promedio, podemos esperar que difiera de cero la diferencia observada en las dos proporciones?

Ya hemos visto en este capítulo que el error estándar de una proporción indica con cuanta precisión puede estimarse el valor en una población a partir de una muestra. Podemos definir el error estándar de la diferencia entre dos proporciones de un modo similar. Teóricamente, si repitiéramos muchas veces el estudio clínico anterior de cáncer de cuello uterino utilizando el mismo número de pacientes en cada grupo, obtendríamos una distribución muestral de las diferencias entre proporciones con una forma similar a la mostrada en la Figura 6.1. La dispersión de la distribución muestral podría resumirse calculando el error estándar. Su aplicación a los datos del ensayo de cáncer de cuello uterino proporciona un error estándar igual al 6,6%. La esencia de la prueba estadística que aplicamos a esta situación es calcular cuántos errores estándar se aleja de cero la diferencia observada de las proporciones. Esto se obtiene como sigue:

$$\text{Valor de la prueba estadística} = \frac{\text{diferencia observada en proporciones} - 0}{\text{error estándar de la diferencia}}$$

En el estudio clínico de cáncer de cuello uterino, la prueba estadística tiene un valor de

$$\frac{0,221 - 0}{0,066} = 3,35$$

La diferencia observada entre los dos tratamientos (22,1%) se separa 3,35 errores estándar del valor cero que es el correspondiente a la hipótesis nula. Un valor tan alto como este es muy poco probable que aparezca al azar, pues ya sabemos que el 95% de las observaciones muestrales de una distribución con 'forma de campana' (es decir, una Distribución Normal) se encontrará dentro de $\pm 1,96$ errores estándar alrededor de su centro.

Cuanto mayor sea el valor de la prueba estadística, menor es la probabilidad P de obtener la diferencia observada si la hipótesis nula es cierta. Po-

demos referirnos a tablas que convierten unos valores estadísticos particulares en los valores correspondientes para P . A continuación se muestra un extracto de una tabla de este tipo, basado en la distribución Normal (forma de campana).

La prueba estadística excede en valor absoluto ^a	0,674	con una probabilidad	0,5 (50%)
"	1,282	"	0,2 (20%)
"	1,645	"	0,1 (10%)
"	1,960	"	0,05 (5%)
"	2,576	"	0,01 (1%)
"	3,291	"	0,001 (0,1%)

^a Valor absoluto significa que se ignoran los signos positivos y negativos; por ejemplo, -1 y $+1$ tienen el mismo valor absoluto de 1.

En el ejemplo del cáncer de cuello uterino, el valor de la prueba estadística es 3,35, incluso mayor que el valor más alto mostrado anteriormente (3,291), de modo que la probabilidad P es menor de 0,001. Es decir, si los tratamientos nuevo y estándar fueran realmente igual de eficaces, la probabilidad de obtener una diferencia tan grande en la supervivencia es menor de uno por mil. Según el uso convencional de la terminología estadística diríamos que la diferencia entre los porcentajes es *'estadísticamente significativa'* al nivel del 0,1%. Por tanto, existe una fuerte evidencia de que el tratamiento nuevo está relacionado con una supervivencia a un año superior a la del tratamiento estándar.

6.2.3 Interpretación de los valores de P

Cox y Snell (1981) indican las siguientes pautas generales para interpretar los valores de P :

Si $P > 0,1$, los resultados se corresponden bastante con la hipótesis nula.

Si $P \approx 0,05$, existe una evidencia moderada contra la hipótesis nula.

Si $P \leq 0,01$, existe una fuerte evidencia contra la hipótesis nula.

Suele considerarse un valor de $P < 0,05$ como indicativo de una evidencia sustancial de que la hipótesis nula no es cierta. (La hipótesis nula se rechaza y los resultados del estudio se declaran *'estadísticamente significativos'* al nivel del 5%).

*No obstante, este énfasis sobre un valor crítico de $P < 0,05$ debería reducirse, ya que 0,05 es un valor de corte arbitrario para decidir si los resultados de un estudio son *'estadísticamente significativos'* o *'no estadísticamente significativos'*. Es preferible indicar directamente el valor de P a indicar si P se encuentra por debajo o por encima de unos valores límite arbitrarios.*

Cuando P es grande, los investigadores suelen presentar un resultado *'no estadísticamente significativo'* y proceder como si hubieran demostrado que

no había efecto. Todo lo que hicieron realmente fue no poder demostrar que lo había. La distinción entre demostrar que no hay efecto y no poder demostrar que lo hay es sutil pero muy importante, ya que la magnitud del valor de P depende de dos parámetros, de la magnitud del efecto observado y del número de observaciones realizadas. Por tanto, un número pequeño de observaciones puede dar lugar a un valor de P grande a pesar del hecho que el efecto real sea grande. A la inversa, con un número grande de observaciones, efectos pequeños, tan pequeños como para ser clínica y epidemiológicamente irrelevantes, pueden alcanzar la significación estadística. Estos aspectos son muy importantes para los investigadores clínicos y epidemiólogos y se considerarán en detalle en capítulos posteriores (13 y 15).

6.2.4 Comparación de otros estimadores muestrales

Aunque hemos introducido la prueba de la significación para un problema en particular (comparación de dos proporciones), este procedimiento puede aplicarse a todos los tipos de estudios comparativos. Por ejemplo, para comparar otros estimadores muestrales (p.e. medias, tasas) o para evaluar de manera más general asociaciones entre variables.

Ejemplo 6.4. Supongamos que un grupo de investigadores realiza un estudio caso-control para investigar la relación entre la historia pasada de infertilidad y el riesgo de desarrollar tumores benignos de ovario. Los resultados de este estudio se presentan en la Tabla 6.4.

	Historia de infertilidad	
	Sí ('expuesto')	No ('no expuesto')
Mujeres con tumores de ovario benignos ('casos')	16	42
Mujeres sanas ('controles')	9	120

$$\text{Odds ratio} = \frac{\text{odds de infertilidad entre los casos}}{\text{odds de infertilidad entre los controles}} = \frac{16/42}{9/120} = 5,08$$

(El cálculo de la *odds ratio* para estudios caso-control se discute en detalle en el Capítulo 9.)

Tabla 6.4.

Distribución de los casos y controles de tumores benignos de ovario según la historia previa de infertilidad: datos hipotéticos.

En el Ejemplo 6.4 la hipótesis nula supone que los dos grupos de mujeres presentan el mismo riesgo de tener tumores benignos de ovario, es decir, que la *odds ratio* real en la población es igual a uno. Tras utilizar una prueba estadística apropiada, como se describe en el Anexo 6.1, los investigadores obtuvieron un valor de P de 0,0003, es decir, si la hipótesis nula fuera cierta, la probabilidad de obtener una *odds ratio* superior o igual a 5,08 sería muy pequeña (inferior a 1 por 1.000). Por tanto, los datos de este estudio proporcionan una evidencia fuerte contra la hipótesis nula.

6.3 Intervalo de confianza y prueba de la hipótesis

Las pruebas de significación estadística están diseñadas para decidir si una serie de observaciones son compatibles o no con algunas hipótesis pero no proporcionan información sobre la magnitud de la asociación (o la diferencia de los efectos). Es más informativo pensar no sólo en términos de prueba de la significación estadística sino también en estimar el tamaño del efecto junto con alguna medida de la incertidumbre en esta estimación.

Este razonamiento no es nuevo; lo utilizamos en el Apartado 6.1 cuando introdujimos los intervalos de confianza. Indicamos que el intervalo de confianza al 95% para un estimador muestral podía calcularse como sigue:

$$\text{estimador muestral} \pm 1,96 \times \text{error estándar del estimador muestral}$$

Un intervalo de confianza al 95% para la diferencia en las dos proporciones puede calcularse de un modo similar:

$$\text{diferencia observada} \pm 1,96 \times \text{error estándar de la diferencia}$$

(el cálculo del EE de la diferencia entre proporciones se ilustra en el Anexo, Apartado A6.1.2).

En el estudio clínico del cáncer de cuello uterino (Ejemplo 6.3), este intervalo de confianza al 95% es

$$0,221 \pm 1,96 \times 0,066 = 0,092 \text{ a } 0,350 = 9,2\% \text{ a } 35,0\%$$

Por tanto, puede considerarse que la diferencia real en la supervivencia a un año entre el tratamiento nuevo y el tratamiento estándar se encuentra entre el 9% y el 35%.

Este intervalo de confianza se corresponde con el resultado de la prueba estadística que realizamos anteriormente (Apartado 6.2.2). El valor de la prueba para la hipótesis nula según la cual no habría ninguna diferencia entre los dos tratamientos fue de 3,35, que se correspondía con un valor de $P < 0,001$.

Observe que si el intervalo de confianza al 95% para una diferencia *no incluye el valor de cero de la hipótesis nula*, entonces el valor de P es menor de 0,05. Por el contrario, si este intervalo de confianza incluye el valor 0, es decir, un límite es positivo y el otro negativo, entonces el valor de P es mayor de 0,05.

Este ejemplo muestra que existe una estrecha relación entre las pruebas de significación y los intervalos de confianza. Esto no es sorprendente ya que estas dos aproximaciones utilizan el mismo componente, el error estándar. Una prueba estadística se basa en calcular a cuántos errores estándar desde el valor especificado en la hipótesis nula está el estimador muestral observado. Los intervalos de confianza utilizan el error estándar para calcular un intervalo de valores que es probable que incluya el parámetro real, pero desconocido de la población; este intervalo puede incluir o no el

valor especificado en la hipótesis nula y esto se refleja por el valor de la prueba estadística.

Ejemplo 6.5. Varios estudios han indicado que los familiares de pacientes que desarrollan micosis fungoide (una forma especial de linfoma no Hodgkin) presentan un mayor riesgo de desarrollar cáncer, especialmente linfomas no Hodgkin. Para clarificar este tema, se evaluaron los datos de un número de casos de cáncer diagnosticados en parientes de primer grado de pacientes con micosis fungoide en un hospital de Londres. El número observado de casos de cáncer se comparó con el número de casos esperado calculado a partir de las tasas nacionales de Inglaterra y Gales. Los resultados de este estudio se presentan en la Tabla 6.5.

Localización	Número de casos observados (O)	Número de casos esperados (E) ^a	Razón de incidencia estandarizada (O/E)	Intervalo confianza al 95%	Valor P
Todas las localizaciones	34	36,8	0,9	0,6–1,3	0,719
Linfoma no Hodgkin	3	2,1	1,5	0,5–4,6	0,502

^a Calculado utilizando como estándar las tasas de incidencia específicas de edad para Inglaterra y Gales.

Tabla 6.5. Incidencia de cáncer entre parientes de pacientes con micosis fungoide: datos no publicados.

En el Ejemplo 6.5, los valores de P son muy altos, indicando que los resultados son consistentes con la hipótesis nula, según la cual no existe ninguna diferencia en el riesgo entre los parientes de pacientes con micosis fungoide y la población general de Inglaterra y Gales. No obstante, la observación de los intervalos de confianza muestra que el intervalo de confianza para todos los tipos de cáncer es bastante estrecho, mientras que el del linfoma no Hodgkin es amplio y compatible tanto con un aumento en el riesgo de casi 5 veces, como con una reducción del riesgo del 50%. Este intervalo de confianza es amplio porque sólo está basado en tres casos.

Para resumir, los valores de P no deben presentarse aislados. Los intervalos de confianza son mucho más informativos ya que proporcionan una idea de la magnitud posible de la asociación (o de la diferencia de efectos) y su amplitud indica el grado de incertidumbre en la estimación del efecto.

En el Anexo 6.1 se presentan fórmulas para calcular los intervalos de confianza y las pruebas estadísticas para la mayoría de las medidas epidemiológicas utilizadas habitualmente.

Lecturas adicionales

* Gardner y Altman (1986) proporcionan una revisión simple de la mayoría de los conceptos tratados en este capítulo y también indican sugerencias para la presentación y representación gráfica de los resultados estadísticos.

Cuadro 6.1. Puntos esenciales

- Los estudios epidemiológicos se realizan normalmente en subgrupos o muestras de individuos tomadas de la población de interés. Por tanto, es poco probable que el estimador muestral de una medida epidemiológica en particular sea igual al valor real de la población debido al *error de muestreo*.
- El *intervalo de confianza* indica la precisión en que es probable que el estimador muestral contenga el valor real del parámetro de interés en la población. Proporciona un rango de valores que es probable que incluya el valor real de la población (aunque no podemos estar seguros de que un intervalo de confianza particular lo haga).
- La *prueba de significación estadística* se utiliza para probar hipótesis sobre el valor de algún parámetro en la población de interés. El valor P proporciona una medida del grado en que los datos del estudio se corresponden con la 'hipótesis nula', es decir, con la hipótesis según la cual no existe una relación real entre variables o diferencia entre grupos en la población. Por tanto, cuanto menor es el valor de P , más fuerte es la evidencia contra la hipótesis nula y, por consiguiente, más fuerte es la evidencia a favor de la hipótesis del estudio.
- Los valores de P en general no deben presentarse aislados ya que no proporcionan ninguna indicación de la magnitud de la asociación (o de la diferencia de efectos). Por ejemplo, pequeños efectos sin relevancia epidemiológica pueden convertirse en 'estadísticamente significativos' si están basados en tamaños de muestra grandes, mientras que efectos importantes pueden ser 'no estadísticamente significativos' debido a que el tamaño de la muestra estudiada sea demasiado pequeño. Por el contrario, los intervalos de confianza proporcionan una idea del intervalo de valores que puede incluir el valor real.
- Los intervalos de confianza y la prueba de la significación estadística tratan sólo la variación del muestreo. Suponen que los errores que no son de muestreo tales como el sesgo en la selección de los sujetos de la muestra y en la determinación de las variables de interés no están presentes.

Anexo 6.1

Intervalos de confianza y pruebas de significación para las medidas epidemiológicas

Este anexo proporciona fórmulas para el cálculo de los intervalos de confianza y las pruebas de la significación estadística para la mayoría de las medidas epidemiológicas utilizadas con frecuencia. Las fórmulas presentadas sólo pueden aplicarse a medidas 'crudas' (a excepción de la Tasa de Mortalidad (o Incidencia) Estandarizada). En el Capítulo 14 se comentan las medidas ajustadas por el efecto de factores de confusión potenciales. En Armitage y Berry (1994) se tratan otras medidas no consideradas aquí. Unos resultados similares pueden obtenerse fácilmente en paquetes informáticos tales como EPI INFO, STATA o EGRET.

A6.1.1 Cálculo de los intervalos de confianza para medidas de aparición de la enfermedad

Proporción única (prevalencia o riesgo)

La prevalencia es una *proporción* y por tanto el error estándar y el intervalo de confianza pueden calcularse utilizando la fórmula comentada en el Apartado 6.1.1:

$$EE(p) = \sqrt{\frac{p^2(1-p)}{a}}$$

donde a es el número de casos y $p = a/n$ (siendo n el tamaño de la muestra). Un intervalo de confianza al 95% puede obtenerse como

$$p \pm 1,96 \times EE(p)$$

Para obtener un intervalo de confianza al 90% el valor 1,96 debe sustituirse por 1,64 y para el intervalo de confianza al 99% por 2,58.

El riesgo también es una *proporción*. Por tanto el error estándar y el intervalo de confianza pueden obtenerse exactamente del mismo modo, siempre que todos los sujetos se hayan seguido durante todo el periodo de riesgo que estamos estudiando. Si los tiempos de seguimiento no son iguales, deben utilizarse las técnicas de tablas de vida o análisis de supervivencia, incluyendo las fórmulas adecuadas del error estándar (véase Capítulo 12).

El método simple para obtener los intervalos de confianza descrito anteriormente se basa en aproximar la distribución muestral a la distribución

'Normal'. Este método 'aproximado' es exacto en muestras suficientemente grandes (mayores de 30).

Para muestras más pequeñas se recomienda un método 'exacto' para calcular los intervalos de confianza para las proporciones, basado en la distribución binomial. No obstante, este método es demasiado complejo para realizarlo con una calculadora.

Tasa simple

Si el número de casos que aparecen durante el periodo de observación se anota como *a* y el número de persona-tiempo a riesgo como *y*, la tasa de incidencia estimada (*r*) es

$$r = a / y$$

Un error estándar 'aproximado' puede calcularse como sigue

$$EE (r) = r / \sqrt{a}$$

El intervalo de confianza al 95% puede obtenerse como

$$r \pm 1,96 \times EE (r)$$

Tabla A6.1.1.
Factores de los límites de confianza al 95% para estimar una variable sometida a la distribución de Poisson.

Número observado en el que se basa la estimación	Factor de límite inferior	Factor de límite superior	Número observado en el que se basa la estimación	Factor de límite inferior	Factor de límite superior	Número observado en el que se basa la estimación	Factor de límite inferior	Factor de límite superior
(a)	(L)	(U)	(a)	(L)	(U)	(a)	(L)	(U)
1	0,025	5,57	21	0,619	1,53	120	0,833	1,200
2	0,121	3,61	22	0,627	1,51	140	0,844	1,184
3	0,206	2,92	23	0,634	1,50	160	0,854	1,171
4	0,272	2,56	24	0,641	1,48	180	0,862	1,160
5	0,324	2,33	25	0,647	1,48	200	0,868	1,151
6	0,367	2,18	26	0,653	1,47	250	0,882	1,134
7	0,401	2,06	27	0,659	1,46	300	0,892	1,121
8	0,431	1,97	28	0,665	1,45	350	0,899	1,112
9	0,458	1,90	29	0,670	1,44	400	0,906	1,104
10	0,480	1,84	30	0,675	1,43	450	0,911	1,098
11	0,499	1,79	35	0,697	1,39	500	0,915	1,093
12	0,517	1,75	40	0,714	1,36	600	0,922	1,084
13	0,532	1,71	45	0,729	1,34	700	0,928	1,078
14	0,546	1,68	50	0,742	1,32	800	0,932	1,072
15	0,560	1,65	60	0,770	1,30	900	0,936	1,068
16	0,572	1,62	70	0,785	1,27	1000	0,939	1,064
17	0,583	1,60	80	0,798	1,25			
18	0,593	1,58	90	0,809	1,24			
19	0,602	1,56	100	0,818	1,22			
20	0,611	1,54						

^a Datos de Haenszel *et al.*, (1962)

Estas fórmulas son adecuadas cuando el número de casos en el numerador de la tasa, a , es 'grande' (mayor de 30). Si el número de casos es menor que éste, entonces los intervalos de confianza 'exactos' basados en la distribución de Poisson pueden obtenerse a partir de la Tabla A6.1.1. Esta tabla proporciona factores por los cuales se multiplica la tasa observada para obtener el límite inferior y superior de un intervalo de confianza al 95%:

Límite inferior = $r \times$ factor del límite inferior (L)

Límite superior = $r \times$ factor del límite superior (U)

Consideremos el siguiente ejemplo. El número total de muertes por cáncer de estómago en hombres de 45-54 años de edad en Egipto durante 1980 fue de 39 en 1.742.000 personas-años (OMS, 1986). Por tanto, utilizando el método 'aproximado':

$$r = 39 / 1.742.000 \text{ personas-años} = 2,24 \text{ por } 100.000 \text{ p-as}$$

$$EE(r) = 2,24 \text{ por } 100.000 \text{ personas-años} / \sqrt{39} = 0,36 \text{ por } 100.000 \text{ p-as}$$

$$IC95\%(r) = 2,24 \pm 1,96 \times 0,36 = 1,53 \text{ a } 2,95 \text{ por } 100.000 \text{ p-as}$$

Al utilizar el método 'exacto', el factor del límite inferior (L) y el factor del límite superior (U) correspondientes a 39 casos se obtienen a partir de la tabla por interpolación entre las filas para 35 y 40 casos.

$$L = 0,697 + \left[(0,714 - 0,697) \times \frac{39 - 35}{40 - 35} \right] = 0,711$$

$$U = 1,39 - \left[(1,39) - 1,36 \right] \times \frac{39 - 35}{40 - 35} = 1,37$$

Así, los límites del intervalo de confianza al 95% son:

$$\text{Límite inferior} = 2,24 \text{ por } 100.000 \text{ personas-años} \times 0,711 = 1,59 \text{ por } 100.000 \text{ p-as}$$

$$\text{Límite superior} = 2,24 \text{ por } 100.000 \text{ personas-años} \times 1,37 = 3,07 \text{ por } 100.000 \text{ p-as}$$

En este ejemplo, los límites 'exactos' y 'aproximados' son relativamente próximos entre sí porque la tasa se basaba en un número de casos lo suficientemente grande. Cuanto mayor sea el número de casos, más próximos son los límites de confianza obtenidos por estos dos métodos.

Consideremos ahora algunos datos de Kuwait. El número total de muertes por cáncer de estómago en hombres de 45-54 años de edad en este país en 1980 fue sólo de 3 en 74.000 personas-años (OMS, 1983). El método 'aproximado' dará:

$$r = 3 / 74.000 \text{ personas-años} = 4,05 \text{ por } 100.000 \text{ p-as}$$

$$EE (r) = 4,05 \text{ por } 100.000 \text{ p-as} / \sqrt{3} = 2,34 \text{ por } 100.000 \text{ p-as}$$

$$IC95\% (r) = 4,05 \pm 1,96 \times 2,34 = -0,54 \text{ a } 8,64 \text{ por } 100.000 \text{ p-as}$$

Este método proporciona un valor negativo para el límite inferior que no tiene sentido ya que las tasas de incidencia y de mortalidad nunca son negativas. Por el método 'exacto', consultando otra vez Tabla A6.1.1, los límites para el intervalo de confianza al 95% son:

$$\text{Límite inferior} = 4,05 \text{ por } 100.000 \text{ p-as} \times 0,206 = 0,83 \text{ por } 100.000 \text{ p-as}$$

$$\text{Límite superior} = 4,05 \text{ por } 100.000 \text{ p-as} \times 2,92 = 11,83 \text{ por } 100.000 \text{ p-as}$$

En este ejemplo, los intervalos de confianza 'exacto' y 'aproximado' son claramente diferentes. Cuando el número de casos es pequeño, menor de unos 30, es preferible utilizar el método 'exacto'.

A6.1.2 Cálculo de los intervalos de confianza para medidas del efecto

Razón de proporciones (razón de prevalencias o razón de riesgos)

Una fórmula para el intervalo de confianza alrededor del estimador de una razón de riesgos debe tener en cuenta el hecho de que la distribución muestral de los valores posibles para la razón de riesgos está muy desplazada a la derecha. El valor mínimo que puede tomar una razón de riesgos es cero y el máximo es infinito. Para hacer la distribución más simétrica es necesario convertir antes las razones de riesgos estimadas en sus logaritmos naturales (denominados ln). Entonces pueden utilizarse fórmulas análogas a las presentadas en el Apartado A6.1.1 para calcular un intervalo de confianza alrededor del valor del logaritmo de la razón de riesgos más que de la propia razón de riesgos.

Consideremos el siguiente ejemplo en el que 1.000 sujetos expuestos y 1.500 sujetos no expuestos se siguieron durante un año. El seguimiento fue completo para cada sujeto. Al final de este periodo, 60 sujetos entre los expuestos y 45 entre los no expuestos desarrollaron la enfermedad de interés

Tabla A6.1.2).

La razón de riesgo (R) y su logaritmo natural pueden calcularse como

$$R = p_1/p_0$$

$$\text{Ln } R = \ln (p_1/p_0)$$

Tabla A6.1.2.

Resultados de un estudio de cohorte en el que los riesgos se calcularon como medidas de la aparición del resultado de interés en cada grupo de estudio: datos hipotéticos

		Exposición		Total
		Sí	No	
Enfermedad	Sí	60 (a)	45 (b)	105 (n ₁)
	No	940 (c)	1.455 (d)	2.395 (n ₀)
Total		1.000 (m ₁)	1.500 (m ₀)	2.500 (M)

Un error estándar 'aproximado' del logaritmo de R puede estimarse por

$$EE (\ln R) = \sqrt{\frac{1}{a} + \frac{1}{b} - \frac{1}{m_1} - \frac{1}{m_0}}$$

Un intervalo de confianza al 95% 'aproximado' para $(\ln R)$ viene dado entonces por $(\ln R) \pm 1,96 EE (\ln R)$ y el intervalo de confianza al 95% para la razón de riesgos (R) se obtiene al tomar antilogaritmos.

Así, en el ejemplo mostrado en Tabla A6.1.2

$$\text{Riesgo en los expuestos } (p_1) = 60 / 1.000 = 0,06$$

$$\text{Riesgo en los no expuestos } (p_0) = 45 / 1.500 = 0,03$$

$$\text{Razón de riesgos (R)} = 0,06 / 0,03 = 2,0$$

$$\ln R = \ln (2,0) = 0,69$$

$$EE (\ln R) = \sqrt{\frac{1}{60} + \frac{1}{45} - \frac{1}{1.000} - \frac{1}{1.500}} = 0,19$$

$$\text{IC 95\% } (\ln R) = 0,69 \pm 1,96 \times 0,19 = 0,318 \text{ a } 1,062$$

El intervalo de confianza al 95% 'aproximado' para la razón de riesgos (R) puede obtenerse tomando antilogaritmos:

$$\text{IC 95\% para (R)} = e^{0,318} \text{ a } e^{1,062} = 1,37 \text{ a } 2,89$$

Una estrategia similar puede aplicarse cuando la medida de interés es una *razón de prevalencias*.

Los métodos 'exactos' deben utilizarse cuando la razón de riesgos o la razón de prevalencias se basa en un número pequeño de casos pero los cálculos son demasiado complejos para mostrarlos aquí.

Diferencia de proporciones (diferencia de prevalencias o diferencia de riesgos)

El error estándar de la diferencia entre dos proporciones p_1 y p_0 , puede estimarse, aproximadamente, como

$$EE (p_1 - p_0) = \sqrt{\frac{p_1^2 (1 - p_1)}{a} + \frac{p_0^2 (1 - p_0)}{b}}$$

Siendo a y b el número de casos en los dos grupos de estudio.

En el ejemplo mostrado en Tabla A6.1.2 ,

$$\text{Diferencia de riesgos } (p_1 - p_0) = 0,06 - 0,03 = 0,03$$

$$EE (p_1 - p_0) = \sqrt{\frac{0,06^2 (1 - 0,06)}{60} + \frac{0,03^2 (1 - 0,03)}{45}} = 0,0087$$

$$\begin{aligned} IC 95\% (p_1 - p_0) &= (p_1 - p_0) \pm 1,96 EE (p_1 - p_0) \\ &= 0,03 \pm 1,96 \times 0,0087 \\ &= 0,013 \text{ a } 0,047 \text{ ó } 1\% \text{ a } 5\% \end{aligned}$$

El intervalo de confianza para una *diferencia en las prevalencias* se calculará del mismo modo.

Razón de tasas

Consideremos los resultados de otro estudio de cohorte hipotético mostrado en la Tabla A6.1.3.

Tabla A6.1.3.

Resultados de un estudio de cohorte en el que las tasas se calcularon como medidas de la aparición del resultado de interés en cada grupo de estudio: datos hipotéticos

	Exposición		Total
	Sí	No	
Casos	60 (a)	45 (b)	105 (n)
Personas-años a riesgo	4.150 (y ₁)	6.500 (y ₀)	10.650 (y)
Tasas por 1.000 personas-años	14,5 (r ₁)	6,9 (r ₀)	9,9 (r)

Al igual que con la razón de riesgos, la razón de tasas sólo puede tomar valores de cero a infinito. Por tanto para obtener un intervalo de confianza para una razón de tasas estimada (RT), es necesario calcular antes su logaritmo natural

$$\ln RT = \ln (r_1/r_0)$$

Un error estándar 'aproximado' del logaritmo de una razón de tasas (RT) puede obtenerse como sigue

$$EE (\ln RT) = \sqrt{(1/a + 1/b)}$$

donde *a* y *b* son los números de casos en el grupo de expuestos y no expuestos.

En este ejemplo, la tasa de incidencia en los expuestos (*r*₁) es igual a 60/4.150 = 14,5 por 1.000 personas-años. La tasa de incidencia en el grupo de los no expuestos (*r*₀) es 45/6.500 = 6,9 por 1.000 personas-años. Por tanto la razón de tasas y su logaritmo son:

$$RT = 14,5 \text{ por } 1.000 \text{ p-as} / 6,9 \text{ por } 1.000 \text{ p-as} = 2,1$$

$$\ln RT = \ln 2,1 = 0,74$$

El error estándar 'aproximado' para el logaritmo de una razón de tasas de

2,1 basado en 60 casos en el grupo de expuestos y 45 casos en el grupo de no expuestos puede calcularse como sigue:

$$EE(\ln RT) = \sqrt{(1/60 + 1/45)} = 0,197$$

El intervalo de confianza al 95% 'aproximado' del logaritmo de la razón de tasas viene dado por:

$$\begin{aligned} IC\ 95\%(\ln RT) &= \ln RT \pm 1,96 EE(\ln RT) \\ &= 0,74 \pm 1,96 \times 0,197 \\ &= 0,35 \text{ a } 1,13 \end{aligned}$$

Podemos por tanto obtener el intervalo de confianza al 95% 'aproximado' para la razón de tasas tomando los antilogaritmos de estos valores:

$$IC\ 95\%(RT) = e^{0,35} \text{ a } e^{1,13} = 1,42 \text{ a } 3,10$$

También existe un método 'exacto' para calcular los intervalos de confianza para razones de tasas basadas en números de casos reducidos pero su discusión está fuera del objetivo de este capítulo (véase Breslow y Day, 1987, pp. 93-95).

Cuando la razón de tasas es una RME (o RIE) es posible calcular un intervalo de confianza al 95% 'exacto' multiplicando la RME (o RIE) observada por los factores de los correspondientes límites superior e inferior exactamente como hicimos para una tasa individual.

Por ejemplo, si el número observado (O) de muertes por leucemia en una determinada localidad es 20 y sólo se esperaban (E) 15 si la localidad tuviera las mismas tasas específicas por edad que todo el país, la RME habría sido igual a 1,33. Los factores de los límites inferior y superior cuando el número de casos observado es 20 se presentan en la Tabla A6.1.1: 0,611 y 1,54, respectivamente. Así,

$$RME = O/E = 20/15 = 1,33$$

$$\begin{aligned} IC\ 95\%(RME) &= 1,33 \times 0,611 \text{ a } 1,33 \times 1,54 \\ &= 0,81 \text{ a } 2,05 \end{aligned}$$

Diferencia de tasas

El error estándar de la diferencia entre las dos tasas estimadas (r_1 y r_0) viene dado por:

$$EE(r_1 - r_0) = \sqrt{\frac{r_1^2}{a} + \frac{r_0^2}{b}}$$

donde a y b se refieren a los números de casos en los dos grupos.

El intervalo de confianza al 95% vendrá dado por:

$$IC\ 95\% (r_1 - r_0) = (r_1 - r_0) \pm 1,96\ EE (r_1 - r_0)$$

Por tanto, en el ejemplo mostrado en la Tabla A6.1.3,

$$r_1 - r_0 = 14,5\ \text{por 1.000 p-as} - 6,9\ \text{por 1.000 p-as} = 7,6\ \text{por 1.000 p-as}$$

$$EE(r_1 - r_0) = \sqrt{\left((0,0145)^2 / 60 + (0,0069)^2 / 45 \right)} = 0,00214$$

$$= 2,14\ \text{por 1.000 p-as}$$

$$IC\ 95\% (r_1 - r_0) = 7,6 \pm 1,96 \times 2,14 = 3,41\ \text{a}\ 11,79\ \text{por 1.000 p-as}$$

Odds ratio

Los datos de un estudio caso-control pueden presentarse en una tabla 2 x 2 como se muestra a continuación:

Tabla A6.1.4.
Resultados de un estudio caso-control: datos hipotéticos.

	Exposición		Total
	Sí	No	
Casos	457 (a)	26 (b)	483 (n ₁)
Controles	362 (c)	85 (d)	447 (n ₀)
Total	819 (m ₁)	111 (m ₀)	930 (N)

Un error estándar 'aproximado' del logaritmo de una *odds ratio* (OR) puede calcularse como

$$EE (\ln OR) = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

En el ejemplo mostrado en la Tabla A6.1.4,

$$OR = \frac{\text{odds de exposición entre los casos}}{\text{odds de exposición entre los controles}} = \frac{457/26}{362/85} = 4,13$$

$$\ln OR = \ln 4,13 = 1,42$$

$$EE (\ln OR) = \sqrt{(1/457 + 1/26 + 1/362 + 1/85)} = 0,23$$

$$IC\ 95\% (\ln OR) = 1,42 \pm 1,96 \times 0,23 = 0,97\ \text{a}\ 1,87$$

Por tanto el intervalo de confianza al 95% 'aproximado' para la *odds ratio* puede obtenerse tomando los antilogaritmos

$$IC\ 95\% (OR) = e^{0,97}\ \text{a}\ e^{1,87} = 2,64\ \text{a}\ 6,49$$

También puede calcularse un intervalo de confianza 'exacto' para muestras pequeñas pero los cálculos son demasiado complejos para efectuarlos con una calculadora.

A.6.1.3 Pruebas estadísticas

Comparación de dos proporciones (prevalencias ó riesgos)

Para contrastar la hipótesis nula de que no existe una diferencia real entre dos proporciones (*riesgos o prevalencias*), los resultados del estudio deben ordenarse en una tabla 2×2 similar a la Tabla A6.1.2. En esta tabla, el número observado (O) de casos entre los expuestos es a . Podemos calcular el valor esperado (E) en la celda a y la *varianza* (V) suponiendo que la hipótesis nula según la cual no hay diferencias entre los dos grupos sea cierta.

$$O = a$$

$$E = m_1 n_1 / N$$

$$V = \frac{n_1 n_0 m_1 m_0}{N^2 (N - 1)}$$

Una prueba estadística especial denominada test de ji-cuadrado (χ^2) puede aplicarse para medir el grado en que difieren los datos observados de los esperados si las dos proporciones fueran iguales, es decir, si la hipótesis nula fuera cierta.

$$\chi^2 = (O - E)^2 / V$$

En epidemiología, esta aplicación del test de χ^2 toma el nombre de *test de Mantel-Haenszel*.

En el ejemplo mostrado en la Tabla A6.1.2,

$$O = 60$$

$$E = 1.000 \times 105 / 2.500 = 42$$

$$V = \frac{105 \times 2.395 \times 1.000 \times 1.500}{2.500^2 \times (2.500 - 1)} = 24,15$$

$$\chi^2 = \frac{(60 - 42)^2}{24.15} = 13,42$$

Valores grandes de χ^2 indican que los datos son inconsistentes con la hipótesis nula y por tanto existe una relación entre la exposición y la enfermedad (o el resultado de interés). El valor de P se obtiene por comparación del valor calculado de χ^2 con las tablas de la distribución de ji-cuadrado.

Para referir el estadístico calculado por el test de χ^2 con estas tablas, es necesario conocer una cantidad denominada los 'grados de libertad' (gl), que considera el número de subgrupos o 'celdas' en la tabla que contribuyen al cálculo. Para tablas 2×2 el número de grados de libertad (gl) es uno.

Si la hipótesis nula es cierta,

El estadístico χ^2 (con 1 gl) es mayor que	0,45	con una probabilidad de	0,5
"	1,32	"	0,25
"	2,71	"	0,1
"	3,84	"	0,05
"	5,02	"	0,025
"	6,63	"	0,01
"	7,88	"	0,005
"	10,83	"	0,001

Por tanto, si el estadístico χ^2 con 1 gl es mayor que 3,84, entonces $P < 0,05$, indicando alguna evidencia de que existe una diferencia real entre las proporciones. Si es mayor que 6,63, entonces $P < 0,01$ y existe una evidencia fuerte de diferencia.

En el ejemplo mostrado en la Tabla A6.1.2, el valor del test de χ^2 fue 13,42, que corresponde a una $P < 0,001$. Por tanto existe una evidencia fuerte de relación. Podemos concluir que la razón de riesgos observada de 2,0 es diferente de 1 y estadísticamente significativa, y por tanto existe una evidencia muy fuerte de que el riesgo es mayor en los expuestos que en los no expuestos.

De hecho, en el Apartado 6.2.2 ya hemos realizado una prueba de la diferencia entre dos proporciones. Entonces utilizamos una prueba estadística diferente que proporciona resultados similares a los del test de Mantel-Haenszel utilizado aquí.

Observe que la prueba estadística es la misma independientemente de la medida del efecto (*razón de riesgos (o prevalencias) o diferencias de riesgos (o prevalencias)*) en que estemos interesados. No obstante, los intervalos de confianza se calculan de modo diferente para cada medida del efecto (véase Apartado A6.1.2).

Comparación de dos odds

El método comentado anteriormente para comparar proporciones también puede utilizarse para probar la hipótesis nula de que no existen diferencias entre la *odds* de la exposición entre los casos y la *odds* de la exposición entre los controles, es decir, la odds ratio es igual a uno.

En el ejemplo mostrado en la Tabla A6.1.4

$$O = 457$$

$$E = \frac{819 \times 483}{930} = 425,35$$

$$V = \frac{483 \times 447 \times 819 \times 111}{930^2 \times (930 - 1)} = 24,43$$

$$\chi^2 = \frac{(457 - 425,35)^2}{24,43} = 41,0$$

La χ^2 proporciona una medida del grado en que los datos observados difieren de los esperados si las dos *odds* de la exposición fueran iguales. Este valor de χ^2 (con un grado de libertad) corresponde a un valor de $P < 0,001$. Por tanto, existen fuertes evidencias en contra de la hipótesis nula.

Comparación de dos tasas

En estudios de cohortes en los que se utilizan tasas más que riesgos como medida de la frecuencia de la enfermedad, debe considerarse la distribución de personas-años entre los grupos de expuestos y no expuestos.

Veamos de nuevo el ejemplo mostrado en la Tabla A6.1.3. El número observado de casos entre los expuestos es $a = 60$. El valor esperado en la celda a y la varianza suponiendo que la hipótesis nula es cierta, es decir, que no existen diferencias reales en las tasas entre los grupos de expuestos y no expuestos, puede calcularse como sigue:

$$E = ny_1/y \text{ y } V = ny_1y_0/y^2$$

Entonces

$$\chi^2 = (O - E)^2 / V$$

En el ejemplo mostrado en la Tabla A6.1.3,

$$O = 60$$

$$E = \frac{105 \times 4.150}{10.650} = 40,92$$

$$V = \frac{105 \times 4.150 \times 6.500}{(10.650)^2} = 24,97$$

$$\chi^2 = \frac{(60 - 40,92)^2}{24,97} = 14,58$$

Este valor de χ^2 con un grado de libertad corresponde a un valor de $P < 0,001$, proporcionando fuertes evidencias contra la hipótesis nula de ninguna asociación entre la exposición y la incidencia de la enfermedad.

El mismo procedimiento puede aplicarse cuando la razón de tasas es una RME. En este caso, la varianza es igual al número esperado de casos. Por tanto, si el número observado de muertes por leucemia (O) en una localidad en particular es 20 y el número esperado basado en las tasas nacionales específicas para la edad (E) es 15 entonces

$$RME = O/E = 20/15 = 1,33$$

$$V = 15$$

$$\chi^2 = \frac{(20 - 15)^2}{15} = 1,67$$

Este valor de χ^2 con un grado de libertad corresponde a un valor de P tal que $0,1 < P < 0,25$. Por tanto estos resultados se corresponden con la hipótesis nula de ausencia de diferencias en las tasas de mortalidad específicas por edad de leucemia entre la localidad y todo el país.

Observe que la prueba estadística es la misma independientemente de la medida del efecto (*razón de tasas o diferencia de tasas*) en que estemos interesados. No obstante, los intervalos de confianza se calculan de modo diferente (véase Apartado A6.1.2).

Prueba χ^2 de una tendencia lineal en proporciones

Hasta ahora sólo hemos considerado situaciones en que los individuos se clasificaban como 'expuestos' o 'no expuestos'. No obstante, en muchas circunstancias, los individuos también pueden clasificarse en diferentes niveles de exposición. Por ejemplo, supongamos que se efectuó un estudio para evaluar si la infección con virus del papiloma humano (VPH) estaba relacionada con el número de parejas sexuales. Los resultados de este estudio hipotético se muestran en la Tabla A6.1.5.

Tabla A6.1.5.
Distribución de mujeres infectadas y no infectadas por el virus del papiloma humano (VPH) en relación al número de parejas sexuales a lo largo de su vida: datos hipotéticos.

	Número de parejas sexuales a lo largo de su vida				Total
	1	2-3	4-5	>5	
VPH-positivas	19(a_0)	33(a_1)	51(a_2)	107(a_3)	210(n_1)
VPH-negativas	71(b_0)	68(b_1)	42(b_2)	61(b_3)	242(n_0)
Total	90(m_0)	101(m_1)	93(m_2)	168(m_3)	452(N)
Porcentaje de VPH positivas	21,1	32,7	54,8	63,7	46,5
Puntuación	0(x_0)	1(x_1)	2(x_2)	3(x_3)	

Los resultados parecen avalar la hipótesis de que existe una tendencia creciente en la proporción de mujeres VPH positivas con el aumento del número de parejas sexuales. Aunque existe una tendencia lineal aparente en las proporciones en la tabla anterior, cada proporción (o porcentaje) está sujeta a la variabilidad del muestreo. Podemos utilizar un test de χ^2 para la tendencia en las proporciones, y evaluar si la tendencia lineal puede ser debida al azar.

El primer paso es asignar una puntuación x a cada categoría de exposición. Por ejemplo, en la Tabla A6.1.5 se asignó «0» a aquellas mujeres con 1 pareja, «1» a aquellas con 2-3 parejas, etc. El segundo paso es utilizar las siguientes fórmulas para obtener los valores de T_1 , T_2 y T_3 . En estas fórmulas, el símbolo Σ significa suma y el subíndice i equivale a 0, 1, 2, 3, etc.

$$T_1 = \sum a_i x_i = (19 \times 0) + (33 \times 1) + (51 \times 2) + (107 \times 3) = 456$$

$$T_2 = \sum m_i x_i = (90 \times 0) + (101 \times 1) + (93 \times 2) + (168 \times 3) = 791$$

$$T_3 = \sum m_i x_i^2 = (90 \times 0^2) + (101 \times 1^2) + (93 \times 2^2) + (168 \times 3^2) = 1.985$$

La prueba χ^2 para tendencia, tiene 1 grado de libertad y puede calcularse como:

$$\chi^2 = \frac{N (NT_1 - n_1 T_2)^2}{n_1 n_0 (NT_3 - T_2^2)}$$

Por tanto, en nuestro ejemplo:

$$\chi^2 = \frac{452 \times (452 \times 456 - 210 \times 791)^2}{210 \times 242 \times (452 \times 1.985 - 791^2)} = 52,41$$

Una χ^2 de 52,41 con 1 gl corresponde a un valor de $P < 0,0001$. Por tanto podemos concluir que existe fuerte evidencia de una tendencia lineal creciente en la proporción de mujeres VPH positivas a medida que aumenta el número de parejas sexuales en la vida.

Prueba χ^2 de tendencia lineal en odds ratios

Consideremos los datos de un estudio caso-control hipotético efectuado para evaluar la existencia de una disminución del riesgo de desarrollar tumores de ovario epiteliales benignos al aumentar la paridad (Tabla A6.1.6).

	Paridad			Total
	0 ^a	1-2	≥3	
Casos de tumores benignos	30 (a_0)	23 (a_1)	7 (a_2)	60 (n_1)
Controles	46 (b_0)	48 (b_1)	35 (b_2)	129 (n_0)
Total	76 (m_0)	71 (m_1)	42 (m_2)	189 (N)
Odds ratio	1	0.73	0.31	
Puntuación	0 (x_0)	1 (x_1)	2 (x_2)	

^a Tomado como la categoría de referencia en el cálculo de las odds ratios

Tabla A6.1.6.

Distribución de los casos de tumores de ovario benignos y controles según la paridad: datos hipotéticos.

Los resultados aparentemente avalan la hipótesis del estudio. (El cálculo de las odds ratios en estudios caso-control se comentará con detalle en el Capítulo 9).

Puede utilizarse una prueba χ^2 de tendencia lineal para contrastar la hipótesis de la existencia de una tendencia decreciente en el riesgo de tumores de ovario benignos al aumentar la paridad. Los cálculos son exactamente los mismos que los utilizados en la prueba χ^2 de tendencia lineal en proporciones.

$$T_1 = \sum a_i x_i = (30 \times 0) + (23 \times 1) + (7 \times 2) = 37$$

$$T_2 = \sum m_i x_i = (76 \times 0) + (71 \times 1) + (42 \times 2) = 155$$

$$T_3 = \sum m_i x_i^2 = (76 \times 0^2) + (71 \times 1^2) + (42 \times 2^2) = 239$$

La prueba χ^2 para tendencia lineal puede calcularse como:

$$\chi^2 = \frac{189 \times (189 \times 37 - 60 \times 155)^2}{60 \times 129 \times (189 \times 239 - 155^2)} = 6,15$$

Este resultado con 1 gl corresponde a un valor de P tal que $0,01 < P < 0,025$. Por lo tanto existe evidencia de que el riesgo de desarrollar un tumor de ovario benigno disminuye al aumentar la paridad.

Prueba χ^2 de tendencia lineal en tasas

Puede utilizarse un método similar para evaluar la tendencia en las tasas. Considere el siguiente estudio de cohorte para contrastar la hipótesis de que el riesgo de cáncer de mama aumenta con la duración del uso de anticonceptivos orales.

Tabla A6.1.7.

Distribución de casos de cáncer de mama y personas-años a riesgo en función de la duración del uso de anticonceptivos orales: datos hipotéticos.

	Duración del uso de anticonceptivos orales			Total
	0	1-2	≥ 3	
Casos de cáncer de mama	62 (a_0)	42 (a_1)	22 (a_2)	126 (n)
Personas-años a riesgo	31.200 (y_0)	25.100 (y_1)	11.600 (y_2)	67.900 (y)
Tasa por 100.000 p-as	198,7 (r_0)	167,3 (r_1)	189,7 (r_2)	185,6 (r)
Puntuación	0 (x_0)	1 (x_1)	2 (x_2)	

Las tasas observadas por nivel de exposición indican, en todo caso, una tendencia decreciente al aumentar la duración del uso de anticonceptivos orales. Un test de χ^2 para tendencia lineal en tasas similar a la descrita anteriormente para las proporciones y *odds ratios* puede calcularse como sigue:

$$T_1 = \sum a_i x_i = (62 \times 0) + (42 \times 1) + (22 \times 2) = 86$$

$$T_2 = \sum y_i x_i = (31.200 \times 0) + (25.100 \times 1) + (11.600 \times 2) = 48.300$$

$$T_3 = \sum y_i x_i^2 = (31.200 \times 0^2) + (25.100 \times 1^2) + (11.600 \times 2^2) = 71.500$$

La prueba χ^2 para tendencia lineal en las tasas, con un grado de libertad, puede calcularse como:

$$\chi^2 = \frac{[T_1 - (n / y)T_2]^2}{(n / y^2)(yT_3 - T_2^2)}$$

Por tanto, en nuestro ejemplo

$$\chi^2 = \frac{[86 - (126 / 67.900 \times 48.300)]^2}{(126 / 67.900^2) \times (67.900 \times 71.500 - 48.300^2)} = 0,19$$

Este resultado con 1 gl corresponde a un valor de $P > 0,5$. Por tanto, los resultados del estudio no confirman la presencia una tendencia creciente o decreciente en las tasas de cáncer de mama con la duración del uso de anticonceptivos orales.

Validez de los pruebas χ^2

Si en una tabla 2×2 el tamaño total de la muestra (N) es menor de 40 y el valor esperado en cualquiera de las celdas es inferior a 5, no debe utilizarse la χ^2 . En estas circunstancias, la prueba estadística apropiada será la Prueba Exacta de Fischer (véase Kirkwood (1988)). Para tablas mayores, la prueba χ^2 es válida, si no hay más del 20% de los valores esperados inferiores a 5, y ninguno es inferior a uno.

El valor esperado (E) para una celda en particular se calcula como sigue:

$$E = \frac{\text{Total de la fila pertinente} \times \text{total de la columna pertinente}}{N}$$

Por tanto, en la Tabla A6.1.4,

$$E(a) = n_1 m_1 / N = (483 \times 819) / 930 = 425,35$$

$$E(b) = n_1 m_0 / N = (483 \times 111) / 930 = 57,65$$

$$E(c) = n_0 m_1 / N = (447 \times 819) / 930 = 393,56$$

$$E(d) = n_0 m_0 / N = (447 \times 111) / 930 = 53,35$$

La prueba χ^2 es válida en este ejemplo porque el tamaño total de la muestra (N) es mayor de 40 y todos los valores esperados de las celdas son superiores a 5.

Capítulo 7

Estudios de intervención

Los estudios de intervención son los diseños epidemiológicos que más se asemejan a los experimentos realizados por los científicos en el laboratorio. El aspecto esencial y distintivo de estos estudios es el control directo del investigador sobre la asignación de los sujetos a los grupos de estudio. Por el contrario, en los estudios observacionales la asignación la determinan los propios sujetos y los investigadores son sólo observadores pasivos de lo que ocurre.

Los estudios de intervención proporcionan una evidencia más sólida cuando se trata de verificar hipótesis. A pesar de esto, no es el tipo de estudio más frecuente en epidemiología, debido principalmente a las limitaciones éticas. Por ejemplo, no sería ético distribuir a las personas en grupos para ser o no expuestas a una sustancia o a un procedimiento que se sospecha puede ser peligroso. No obstante, puede realizarse un ensayo para verificar si la eliminación de esta exposición disminuye la incidencia y mortalidad. Así, los estudios de intervención en epidemiología están limitados a intervenciones con buenas bases teóricas para esperar que aporten un beneficio a los participantes.

7.1 Tipos de estudios de intervención

Los estudios de intervención pretenden prevenir la enfermedad (estudios de campo) o tratar los procesos patológicos establecidos (ensayos clínicos).

El objetivo de un *ensayo clínico* es evaluar uno o más tratamientos nuevos para una enfermedad u otro evento de interés. Por ejemplo, un ensayo clínico puede diseñarse para evaluar si un agente quimioterápico previene la recidiva de cáncer, aumenta la supervivencia o mejora la calidad de vida (Ejemplo 7.1). Debido a que los ensayos clínicos implican la mayoría de las veces a personas enfermas, se realizan en hospitales u otros centros asistenciales donde los sujetos son habitualmente tratados y seguidos por su enfermedad.

Ejemplo 7.1. Un total de 474 pacientes adultos con glioma maligno (astrocitoma) de grado 3 ó 4 se distribuyeron aleatoriamente para recibir 45 Gy (en 20 fracciones durante 4 semanas) o 60 Gy (en 30 fracciones durante 6 semanas) de radioterapia postoperatoria. El principal objetivo del estudio era evaluar si la dosis más alta mejoraría la supervivencia (Bleehen y Stenning, 1991).

Por el contrario, los *estudios de campo* tratan con sujetos no enfermos. Un ensayo de campo evalúa si un agente o procedimiento reduce el riesgo de desarrollar la enfermedad entre las personas exentas de esta condición en el momento del reclutamiento. Debido a que estos ensayos incluyen a personas sanas en lugar de personas enfermas, tienden a ser logísticamente más difíciles de llevar a cabo que los ensayos clínicos. Generalmente tienen que realizarse «sobre el terreno» más que en hospitales o clínicas. Además, mientras las consecuencias adversas (p.e. recidiva de la enfermedad, muerte) de una determinada enfermedad pueden producirse con gran probabilidad durante un tiempo relativamente corto, normalmente el riesgo de contraer la enfermedad entre personas que inicialmente no la tienen es bajo. Esto es especialmente cierto para enfermedades raras tales como el cáncer. Por consiguiente, los estudios de campo habitualmente requieren el seguimiento de un número mayor de sujetos durante periodos más prolongados que los ensayos clínicos.

Ejemplo 7.2. *Un estudio aleatorizado se realizó entre empleados públicos de Whitehall (ingleses) para medir, en hombres de edad media, los efectos de dejar el hábito de fumar sobre la salud. Un total de 1.445 hombres fumadores de cigarrillos, de 40 a 59 años, con un alto riesgo de desarrollar enfermedades cardiorrespiratorias, se distribuyeron aleatoriamente a un grupo de intervención (714 hombres) o a un grupo sin intervención, con asistencia sanitaria normal (731 hombres). Los del grupo de intervención recibieron consejo individual sobre la relación entre fumar y la salud. La mayoría solicitó dejar de fumar y recibió más apoyo durante los siguientes 12 meses. Los dos grupos se siguieron durante veinte años (Rose y Colwell, 1992).*

Los ensayos de campo pueden efectuarse entre *individuos* (como en el Ejemplo 7.2) o en *grupos de personas* (como en los Ejemplos 7.3 y 7.4). En el primer caso, la unidad de asignación a la intervención es el individuo, mientras que en el segundo es el grupo. El grupo puede ser una unidad familiar, un bloque de casas, un colegio o toda una comunidad. Los ensayos de campo en los que comunidades enteras son la unidad de asignación se denominan *ensayos comunitarios* (Ejemplo 7.3).

Ejemplo 7.3. *El Estudio Comunitario de Intervención para Dejar de Fumar [The Community Intervention Trial for Smoking Cessation (COMMIT)] es un proyecto multicéntrico diseñado para evaluar un programa comunitario antitabaco en los EE.UU. Este estudio en curso empezó en 1989 en 11 pares de comunidades con características semejantes entre los pares. Una comunidad de cada par se asignó aleatoriamente para recibir el programa antitabaco, mientras que la otra actuaba como control. La intervención se diseñó para favorecer el dejar de fumar utilizando una gran cantidad de recursos comunitarios para modificar la actitud de la comunidad en relación al tabaco (COMMIT Research Group, 1991).*

Ejemplo 7.4. *Un ensayo aleatorizado controlado se realizó para evaluar la eficacia de los folletos de educación sanitaria en la reducción de la incidencia de quemaduras por el sol (uno de los factores de riesgo conocidos de melanoma maligno cutáneo) en turistas británicos. La población de estudio estaba formada por británicos que iban de vacaciones a países más cálidos en vuelos desde el aeropuerto de Manchester, en el Reino Unido, durante el mes de agosto de 1993. La unidad de estudio fue el vuelo. Estos se distribuyeron aleatoriamente para recibir (grupo de intervención) o no (grupo control) los folletos informativos (Dey et al., 1995).*

Existen varias razones para elegir grupos mejor que individuos como unidad de estudio. Muchas intervenciones son imposibles de asignar individualmente, como las medioambientales: la fluoración del agua o la mejora de la calidad del aire sólo pueden realizarse en grupos. La mayoría de las intervenciones sanitarias educativas también se encuentran en esta categoría. Por ejemplo, la intervención en el Ejemplo 7.3 (es decir, el programa de dejar de fumar) estaba destinada principalmente a la comunidad más que al individuo; por tanto, era más adecuado elegir la comunidad como unidad de estudio. También puede ser logísticamente más fácil realizar el ensayo entre grupos de personas que entre individuos. En el Ejemplo 7.4, era mucho más fácil asignar los vuelos a un grupo de intervención o a un grupo control de lo que habría sido asignar individuos. Mediante la asignación de vuelos se redujo mínimamente el potencial de 'contaminación', es decir, la posibilidad de que personas del grupo control acabaran teniendo acceso a los folletos. Esta 'contaminación' habría hecho que los dos grupos se parecieran más y, como consecuencia, habría disminuido la capacidad del ensayo para detectar el efecto real que pudiera tener la intervención (Apartado 7.10).

7.2 Formulación de los objetivos del estudio

Los principales objetivos del estudio de intervención deben especificarse claramente antes de su inicio. Deben incluir una descripción concisa, pero detallada, de la intervención a evaluar, el resultado(s) de interés y la población en la que se realizará el estudio. Por ejemplo, no es suficiente indicar sólo que el objetivo de un ensayo es 'evaluar si la administración de tamoxifeno previene el desarrollo de cáncer de mama en mujeres'. Es necesario definir exactamente cuál es la población diana. Por ejemplo, ¿incluye a todas las mujeres o sólo aquellas que presentan un elevado riesgo de desarrollar la enfermedad? ¿Qué grupos de edad se incluirán? La intervención tiene que ser especificada en términos de dosis, frecuencia de administración y duración. Además, será necesario decidir si al grupo control se le administrará un placebo o nada. También deben indicarse claramente los resultados de interés y los procedimientos utilizados para medirlos.

También es importante decidir en esta etapa si la intención del estudio es principalmente *científica* (explicatoria) o *pragmática* (Schwartz y Lellouch, 1967). Si es principalmente científica, el ensayo debe efectuarse en condiciones ideales de modo que sea posible establecer cuál es el beneficio máximo que la intervención puede conseguir. En estas circunstancias es sensato realizar el ensayo en grupos especiales de personas, como por ejemplo voluntarios, para así garantizar un alto nivel de cumplimiento. Los ensayos pragmáticos, por el contrario, evalúan si la intervención es eficaz cuando se introduce en una situación de salud pública o clínica, es decir, en condiciones reales. En estos estudios, es posible que el efecto real de la intervención se diluya, entre otras cosas, por los bajos niveles de cumplimiento.

7.3 Aspectos éticos

En los estudios observacionales, es responsabilidad del investigador mantener la confidencialidad de los datos proporcionados por los sujetos del estudio y asegurar que los procedimientos utilizados para medir las exposiciones y los resultados de interés no incluyan niveles inaceptables de incomodidad, estrés o riesgos para los participantes.

En los estudios de intervención, la situación es diferente. Los investigadores ya no observan simplemente lo que ocurre a los sujetos del estudio, debido a que el investigador interviene deliberadamente. Las consideraciones éticas son más importantes en estudios de intervención que en cualquier otro tipo de estudio epidemiológico y sólo están éticamente justificados en una situación de incertidumbre, cuando existe una auténtica duda sobre el valor de una nueva intervención en cuanto a sus ventajas y sus riesgos. Por supuesto, el investigador debe tener una sospecha de que la intervención podría ser beneficiosa, a partir por ejemplo, de estudios de laboratorio y estudios en animales, o a partir de estudios epidemiológicos observacionales. De lo contrario, no habría justificación alguna para realizar un ensayo.

Lamentablemente, muchas intervenciones médicas no se han evaluado correctamente en estudios de intervención bien realizados. Por ejemplo, la mastectomía radical se utilizó durante más de cien años como la forma estándar de tratamiento para las formas tempranas de cáncer de mama. No fue hasta finales de los años 70, cuando finalmente se realizaron ensayos clínicos, en que esta forma de tratamiento se sustituyó por otras más conservadoras. Los ensayos clínicos demostraron que no había diferencias en la recidiva o supervivencia entre la mastectomía radical y otros tipos de cirugía (más conservadores) (Veronesi *et al.*, 1981; Fischer *et al.*, 198x). Por tanto, las mujeres con cáncer de mama en estadios precoces se sometieron de manera innecesaria a una cirugía mutilante durante décadas porque los clínicos estaban convencidos de que no habría sido ético privar a las mujeres del tratamiento estándar. La lección de éste y otros muchos ejemplos es que, por razones éticas y de viabilidad, es mejor realizar un ensayo cuando un agente o procedimiento se introduce por primera vez que cuando ya ha sido ampliamente aceptado y se convierte en práctica estándar. Por tanto, no realizar un

ensayo adecuado, cuando es necesario y factible, también puede ser no ético.

La consideración de si un estudio es ético o no es un juicio subjetivo basado en normas culturales que varían de sociedad a sociedad y a lo largo del tiempo. Una referencia útil con pautas propuestas para investigaciones que impliquen a sujetos humanos son las publicadas por el *Council for International Organizations of Medical Sciences* (CIOMS) y la Organización Mundial de la Salud (OMS) (1993).

7.4 Población diana y población experimental

La *población diana* es el grupo general al cual los investigadores esperan extrapolar los resultados del ensayo. Un ensayo puede afectar a todos los seres humanos si se cree que la intervención a evaluar presenta un beneficio potencial para todos, o sólo a algunos subgrupos de la población, tales como las mujeres o los fumadores. Así, la población diana representa la dimensión del posible impacto de la intervención en términos de salud pública.

Una vez definida la población diana, es necesario elegir la población real en la que se realizará el estudio. La elección de esta *población experimental* depende de una serie de aspectos. En primer lugar, no debe diferir de la población diana de tal modo que no sea posible generalizar los resultados a esta última. En algunas circunstancias, los estudios de intervención se efectúan en grupos especiales, como voluntarios, para garantizar el buen cumplimiento del protocolo o facilitar la logística. Estos ensayos son útiles para evaluar el efecto potencial de una nueva intervención, aunque puede ser difícil extrapolar sus resultados a la población diana. En segundo lugar, es esencial determinar si la población experimental propuesta es lo suficientemente grande como para conseguir el tamaño de muestra necesario para el ensayo (Capítulo 15). En tercer lugar, es importante elegir una población experimental en la que se produzca un número suficiente de eventos de interés para poder realizar comparaciones significativas entre los diferentes tratamientos o procedimientos en un periodo de tiempo razonable. Así, la mayoría de los ensayos se efectúa en poblaciones en las que el riesgo de desarrollar el resultado de interés es alto. Por ejemplo, para evaluar el beneficio potencial

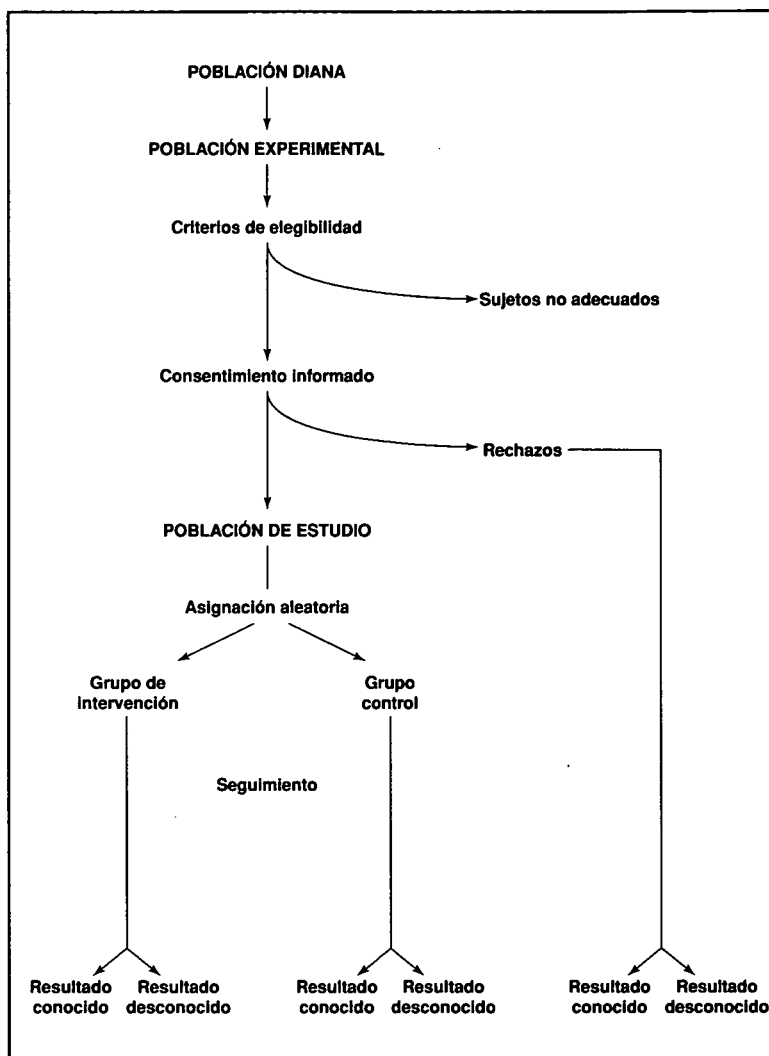


Figura 7.1. Diagrama general de un estudio de intervención.

de un programa para dejar de fumar es lógico elegir una población experimental con una alta prevalencia de fumadores y con una alta incidencia de cáncer de pulmón.

La selección de la población experimental depende también de los factores logísticos. El estudio debe efectuarse en un área donde sea posible obtener apoyo de las autoridades locales o líderes de la comunidad y donde sea posible realizar el seguimiento de los participantes y obtener información completa y exacta durante el desarrollo del ensayo. Por ejemplo, un estudio a largo plazo realizado sobre una población con mucha movilidad como estudiantes universitarios o poblaciones nómadas puede concluir con tasas de seguimiento muy bajas que comprometerían el estudio.

7.5 Criterios de inclusión

Los criterios de inclusión deben establecerse claramente antes del inicio del estudio (Figura 7.1) y especificar exactamente quién debe incluirse en el estudio. Los criterios de inclusión variarán de estudio a estudio según sus objetivos, pero en general, tienden a eliminar sujetos que pueden estar expuestos a un riesgo mayor o que presentan situaciones clínicas subyacentes que puedan interferir en la evaluación. Por ejemplo, los pacientes pueden excluirse si su condición física y/o mental no es adecuada para realizar la entrevista o para la obtención de muestras biológicas. También es habitual excluir mujeres embarazadas o en edad fértil si existe una posibilidad, aunque mínima, de que la intervención pueda dañar al feto.

Una vez definidos los criterios de selección, puede establecerse quiénes son los individuos de la población experimental que son elegibles. A veces es necesario realizar una encuesta basal para identificar los individuos elegibles, como en el Ejemplo 7.5. En estas circunstancias, debe invitarse a los sujetos a participar en la encuesta basal aunque puede que no cumplan los criterios de selección para la inclusión en el estudio.

Ejemplo 7.5. *Se efectuó un estudio de intervención en Gambia para determinar la contribución de los chinches en la transmisión de la hepatitis B. Para ser incluidos en el ensayo, los niños tenían que estar libres de infección por hepatitis B en el momento de la inclusión. Todos los niños con edades comprendidas entre seis meses y cinco años que vivían en siete pueblos adyacentes a Mandika fueron examinados por un médico y se les tomó una muestra de sangre para realizar la serología. Sólo aquellos que no estaban infectados se aleatorizaron al grupo de intervención (pulverización de insecticida de la vivienda del niño) o al grupo control (Mayans et al., 1994).*

7.6 Consentimiento informado

Debe invitarse a los sujetos elegibles a participar en el ensayo. En esta etapa, se les debe informar en un lenguaje sencillo sobre los objetivos y los

procedimientos del estudio, lo que se les pedirá exactamente y los posibles riesgos y beneficios. También se les debe informar que serán asignados al grupo de intervención o al grupo de comparación y que no sabrán a qué grupo se les ha asignado hasta el final del estudio. También debe garantizarse a los sujetos que se respetará su intimidad, que no se revelará su identidad a nadie que no forme parte del equipo de investigación y que los investigadores no utilizarán la información obtenida durante el estudio en su detrimento (por ejemplo, para crear listas fiscales). Debe darse el tiempo suficiente a los individuos para considerar si quieren participar y se les debe permitir negarse a participar o retirarse del estudio en cualquier momento sin consecuencias negativas para ellos.

Si los sujetos a los que se ha proporcionado esta información deciden participar en el estudio, se considera que han dado su *consentimiento informado*. En muchos países, los comités éticos y las autoridades requieren que los participantes firmen un formulario de consentimiento (a veces en presencia de un testigo). Esto puede ser difícil de cumplir y de relativa poca importancia en poblaciones con niveles bajos de alfabetización.

Deben realizarse todos los esfuerzos posibles para intentar explicar la naturaleza del estudio de tal manera que los individuos puedan entenderlo y que sea apropiada para sus valores y pautas culturales. Lo que es adecuado en el mundo 'occidental' no es necesariamente adecuado en otras situaciones culturales. Por ejemplo, en algunas sociedades las decisiones sobre la participación en un estudio en particular pueden tomarse a nivel comunitario más que individualmente, de modo que la autorización para realizar un proyecto de investigación puede obtenerse a través de los dirigentes más respetados de la comunidad, en lugar de sus miembros individuales. Pero incluso aunque se obtenga el consentimiento comunitario para el estudio, el investigador tiene todavía la responsabilidad de explicar el procedimiento del estudio y los riesgos y beneficios potenciales a cada individuo que pueda participar y asegurar que el individuo sabe que es libre de negarse a participar o retirarse de la investigación en cualquier momento.

En las sociedades occidentales existe un paralelismo a esta situación. El consentimiento comunitario antes que el individual se obtiene generalmente en estudios comunitarios ya que suele ser imposible obtener el consentimiento de cada miembro de las comunidades implicadas. El consentimiento debe obtenerse de las autoridades locales y de los dirigentes de la comunidad. Una vez estas personas han aceptado la participación de la comunidad, es importante para el investigador informar a los miembros de la comunidad de que van a participar en un estudio.

7.7 Población de estudio

La *población de estudio* la componen las personas que son elegibles y están dispuestas a participar en el estudio. Con frecuencia son un subgrupo relativamente pequeño y escogido de la población experimental. Es muy probable que los participantes en un estudio de intervención difieran de los no participantes, de manera que puedan afectar el riesgo de desarrollar

la enfermedad en estudio. Que el subgrupo de participantes sea o no representativo de toda la población experimental no afectará a la validez de los resultados de un ensayo realizado en este grupo. No obstante, sí puede afectar a la generalización de los resultados a la población experimental o a la población diana.

Ejemplo 7.6. *El Estudio sobre la Salud de los Médicos [Physicians' Health Study] es un ensayo clínico a doble ciego, aleatorizado, controlado por placebo, realizado en los Estados Unidos para evaluar los efectos de la aspirina sobre la mortalidad total cardiovascular, y del beta-caroteno sobre la incidencia de cáncer. En 1982, se enviaron cartas a 261.248 médicos varones de los EE.UU. entre 40-48 años solicitándoles participar. Casi la mitad respondieron, y de éstos, la mitad estaban dispuestos a participar. Los hombres con historia de cáncer (excepto cáncer de piel no melanoma), infarto de miocardio, apoplejía o isquemia cerebral transitoria se consideraron no elegibles. Por tanto, sólo se encontraron 33.211 médicos dispuestos y elegibles que se incluyeron en la fase de reclutamiento del ensayo, la cual duró de 1 a 6 meses, y en la que los sujetos fueron asignados a recibir aspirina, beta-caroteno o placebo. El objetivo de esta fase de inclusión fue mejorar el cumplimiento ya que sólo se aleatorizaron aquellos médicos que toleraron la aspirina y cumplieron con el régimen de la medicación. Al final de esta fase de inclusión, 22.071 se consideraron elegibles para el ensayo y fueron aleatorizados (Hennekens et al., 1985, 1996; Stampfer et al., 1985).*

En el Ejemplo 7.6, menos de un 10% de la población experimental original se incluyó finalmente en el ensayo. Sólo se aleatorizaron aquellos médicos que demostraron ser excelentes cumplidores y que no experimentaron reacciones adversas, con el fin de aumentar la capacidad (poder) del estudio, para probar las dos hipótesis del proyecto (Capítulo 15). Aunque la exclusión de los poco cumplidores limitaba la posibilidad de generalizar los resultados del ensayo, esto no afectó su validez. En este ejemplo, era mucho más importante obtener respuestas claras a las preguntas planteadas que intentar asegurar que sería posible extrapolar los resultados a una población mayor.

No obstante, debe intentarse obtener datos basales y/o determinar el resultado de la enfermedad de interés en los sujetos elegibles pero no dispuestos a participar. Esta información es extremadamente valiosa para estimar la presencia y la magnitud de las diferencias entre los participantes y los no participantes en un ensayo determinado. Esto ayudará a concluir si los resultados obtenidos a partir de los participantes en el estudio pueden generalizarse a la población diana.

7.8 Elección de la intervención de comparación

Una característica clave de un estudio de intervención es la inclusión de, como mínimo, un grupo control, frente al que se compara el efecto de la

intervención en estudio. En este momento, debe considerarse el tipo de intervención que debe recibir el grupo control. Por ejemplo, ¿en un ensayo clínico, el grupo control debe recibir un placebo (un preparado que se parece al nuevo tratamiento en todos los aspectos a excepción de que no contiene el principio activo), el mejor tratamiento actual o nada?

Si ya existe un tratamiento establecido de valor demostrado, no será ético utilizar un placebo. Además, en estas circunstancias, la cuestión pragmática real no es tanto mostrar si el tratamiento nuevo funciona sino si es mejor que el ya existente. Si no existe un tratamiento estándar, un placebo es justificable porque permite que el estudio sea a doble ciego (véase Apartado 7.11). No obstante, para muchas intervenciones no es posible idear un placebo adecuado. Por ejemplo, no es posible encontrar un placebo adecuado para intervenciones quirúrgicas o para la mayoría de los programas de educación sanitaria.

7.9 Asignación a los diferentes grupos de estudio

Debido a que los participantes y los no participantes pueden diferir en aspectos importantes relacionados con el resultado final, la distribución en los diferentes grupos de estudio debe realizarse *sólo una vez se haya determinado que los sujetos son elegibles y hayan expresado su disposición a participar*. Esto es, los no participantes deben eliminarse del conjunto de sujetos potenciales antes de realizar la distribución en los grupos de intervención y control.

7.9.1 Razones para la distribución aleatoria

La aleatorización es el mejor método para distribuir los sujetos en los diferentes grupos de estudio. Este método permite determinar al azar, y sólo al azar, la asignación de los sujetos a los diferentes grupos. Por tanto, es la única manera de asegurar que cualquier diferencia en los resultados del ensayo es debida a los efectos de la intervención y no a diferencias subyacentes entre los grupos. La aleatorización presenta dos *ventajas principales* en relación a otros métodos de distribución:

- (1) *La aleatorización elimina el sesgo de selección por parte de los participantes e investigadores.*

La aleatorización elimina la posibilidad de toda influencia subjetiva en la asignación de los individuos a los diferentes grupos de estudio. En algunos ensayos también se han utilizado métodos basados en la fecha de nacimiento o fecha de inclusión, asignando a una intervención a los nacidos (o incluidos) en fechas pares y a la otra a los nacidos (o incluidos) en fechas impares. El problema de estos métodos es que el investigador puede saber con anterioridad el grupo al que se asignará al participante y esto puede introducir un sesgo consciente o inconsciente en el procedimiento de distribución. Si el investigador sabe que un sujeto en particular se va a asignar a una intervención concreta, puede estar más o menos dispuesto a conside-

rar al sujeto elegible para la inclusión en el estudio. La aleatorización puede garantizar que esto no ocurra siempre que se efectúe sólo después de verificar que los sujetos sean elegibles y que hayan expresado su voluntad de participar en el ensayo.

- 2) *La aleatorización tiende a crear grupos comparables en términos de distribución de factores conocidos y, más importante, de factores desconocidos que puedan influir en el resultado de interés.*

La aleatorización asegura que la distribución de las variables de confusión conocidas y, más importante, las *desconocidas* sea similar en los grupos a comparar siempre que el tamaño de la muestra sea relativamente grande. Esta es una característica única de los estudios experimentales. En estudios observacionales puede tenerse en cuenta el efecto de las variables de confusión en el momento del análisis. Ahora bien, este ajuste sólo puede realizarse para variables de las que ya se sabía o se sospechaba que lo eran al inicio del estudio y para las cuales se obtuvieron datos (Capítulos 13 y 14). Los ensayos pueden prolongarse durante muchos años y es posible que se conozcan nuevas variables de confusión en este tiempo. No obstante, esto no afectaría a la validez de los resultados del ensayo, ya que la aleatorización habría asegurado que la distribución de las variables de confusión desconocidas fuera similar en los grupos de estudio *siempre que el número de sujetos distribuidos aleatoriamente fuera grande.*

En esta discusión sobre la aleatorización, cabe mencionar que con frecuencia existe una confusión en el uso de expresiones «*distribución aleatoria*» (o «*asignación aleatoria*») y «*muestreo aleatorio*» (o «*selección aleatoria*»). En este apartado estamos tratando la «*distribución aleatoria*» como el proceso por el cual los sujetos se distribuyen en los grupos de estudio en un ensayo. Esto constituye un principio fundamental sobre el que se basan los estudios de intervención. La selección aleatoria se refiere al proceso por el cual se elige al azar una muestra de sujetos de una población más grande. Los ensayos clínicos raramente ocasionan la selección aleatoria; el investigador toma los pacientes disponibles siempre que cumplan los criterios de inclusión en el estudio. En ensayos de campo, puede utilizarse la selección aleatoria si la población experimental es mayor de lo necesario para asegurar que el estudio pueda responder al problema tratado (esto es, que el ensayo tenga un poder o precisión adecuados – véase Capítulo 15). Los métodos para elegir muestras aleatorias de una población se comentarán en el Capítulo 10.

7.9.2 Métodos de aleatorización

Pueden utilizarse varios métodos para distribuir aleatoriamente los sujetos del estudio en los diferentes grupos. Independientemente del método elegido, es importante asegurar desde las etapas iniciales del ensayo que el método de aleatorización y la lista de aleatorización sea inaccesible para las personas responsables de reclutar sujetos, monitorizar los efectos de la intervención y evaluar los resultados del ensayo.

Aleatorización simple

La aleatorización simple es el método más elemental. Es el equivalente a tirar una moneda. No obstante, no debe utilizarse la aleatorización tirando una moneda porque no puede comprobarse o reproducirse. La alternativa es utilizar una tabla de números aleatorios (o una lista de aleatorización generada por ordenador) (Figura 7.2).

El primer paso para determinar asignaciones a grupos aleatorios es establecer una correspondencia entre los números de la tabla y los grupos de estudio. Supongamos que los números impares corresponden al grupo control y los números pares a la nueva intervención. El segundo paso es definir una forma adecuada para leer la tabla de números aleatorios, por ejemplo, leerla verticalmente por columnas o bien horizontalmente por filas.

El tercer paso es elegir un punto de inicio, por ejemplo, cerrando los ojos y eligiendo un número con el dedo. Una vez establecido el punto de inicio, los números se leen en la tabla siguiendo la secuencia definida en el paso dos. La Figura 7.2 reproduce parte de una tabla de números aleatorios (en el Anexo 7.1 se encuentra la tabla completa). Supongamos que el punto de inicio elegido es el marcado con un círculo en la tabla y que se ha decidido que los números deben leerse por columnas en dirección descendente. Los 10 primeros números serían 8,9,3,5,7,5,5,9,1,0. El cuarto paso es hacer las asignaciones del tratamiento según el sistema definido anteriormente (Tabla 7.1).

67	19	00	71	74	60	47	21	29	68	02	02	37	03	31
02	94	37	34	02	76	70	90	30	86	38	45	94	30	38
79	78	45	04	91	16	92	53	56	16	02	75	50	95	98
87	75	66	81	41	40	01	74	91	62	48	51	84	08	32
34	86	82	53	91	00	52	43	48	85	27	55	26	89	62
11	05	65	09	68	76	Ⓢ	20	37	90	57	16	00	11	66
52	27	41	14	86	22	98	12	22	08	07	52	74	95	80
07	60	62	93	55	59	33	82	43	90	49	37	38	44	59
04	02	33	31	08	39	54	16	49	36	47	95	93	13	30
01	90	10	75	06	40	78	78	89	62	02	67	74	17	33
92	03	51	59	77	59	56	78	06	83	52	91	05	70	74
61	71	62	99	15	06	51	29	16	93	58	05	77	09	51
73	32	08	11	12	44	95	92	63	16	29	56	24	29	48
42	10	50	67	42	32	17	55	85	74	94	44	67	16	94
26	78	63	06	55	13	08	27	01	50	15	29	39	39	43

Figura 7.2.

Extracto de una tabla de números aleatorios.

Paciente nº	Sexo	Nº aleatorio	Asignación
1	M	8	Grupo de intervención
2	H	9	Grupo control
3	H	3	Grupo control
4	H	5	Grupo control
5	H	7	Grupo control
6	M	5	Grupo control
7	M	5	Grupo control
8	M	9	Grupo control
9	M	1	Grupo control
10	M	0	Grupo de intervención

Tabla 7.1.

Ejemplo que ilustra el uso de una tabla de números aleatorios para asignar diez sujetos a los grupos del estudio (véase texto).

Las tablas de números aleatorios se generan de tal modo que cada uno de los dígitos de cero a nueve tiene la misma probabilidad de aparecer. Si son necesarios números iguales de participantes en cada grupo de intervención, debe asignarse la misma cantidad de números de un dígito a cada grupo, incluso aunque esto signifique que algunos dígitos no correspondan a ningún grupo. Por tanto, en el caso de tres grupos, tres de los diez números de un dígito se asignan a cada grupo (p.e. los números 1, 2, 3 al grupo A; 4, 5, 6 al grupo B; y 7, 8, 9 al grupo C). El otro número (es decir, cero en nuestro ejemplo) se ignora en las tablas aleatorias y la selección se desplaza al siguiente número.

Uno de los inconvenientes de la aleatorización simple es que por azar puede dar lugar a números de sujetos asignados a cada grupo considerablemente desiguales. Por ejemplo, en el ejemplo anterior sólo dos personas de diez se asignaron al grupo de intervención. Además, la aleatorización simple también puede ocasionar que los diferentes grupos de intervención sean diferentes respecto a factores que puedan afectar al resultado del estudio. En el ejemplo anterior, no sólo era pequeño el número de personas asignadas a la intervención sino que la distribución de sexo también fue bastante diferente en los dos grupos. Esto es probable que se produzca cuando el número total de sujetos en un estudio es bajo. Para ensayos que incluyan varios cientos de participantes, un desequilibrio de este tipo puede ser pequeño y puede tenerse en cuenta a la hora del análisis. Si el número de casos es pequeño, el desequilibrio entre variables puede hacer que el ensayo sea más difícil de interpretar y, por lo tanto, se recomienda asegurar el equilibrio utilizando los métodos de aleatorización descritos a continuación.

Aleatorización restringida (o aleatorización por bloques)

Este método garantiza que el número de participantes asignados a cada grupo de estudio sea igual cada vez que un bloque de un cierto número de pacientes se incluya en el estudio. Supongamos que los pacientes se asignan a los tratamientos A y B de tal modo que cada cuatro sujetos haya el mismo número en cada tratamiento. Sólo existen seis combinaciones posibles (permutaciones) de A y B en bloques de cuatro:

Nº	Combinación
1	AABB
2	ABAB
3	ABBA
4	BBAA
5	BABA
6	BAAB

La combinación para un bloque determinado de 4 pacientes se elige al azar (utilizando una tabla de números aleatorios como se ha descrito anteriormente) de los seis posibles (observese que en el ejemplo anterior los dígitos 7,8,9 y 0 de la tabla de números aleatorios deben ignorarse). Por ejemplo, si los números aleatorios de la tabla fueran 2,3,6,5 (y los bloques se asignaran como se presentan) esto significaría que los pacientes 1-4, 5-8, 9-12, 13-16 recibirían tratamientos ABAB, ABBA, BAAB y BABA, respectivamente. Por consiguiente, ocho pacientes se asignarían al grupo A y otros ocho al grupo B.

Aleatorización estratificada

Cuando los resultados del ensayo es probable que sean distintos en relación al sexo y diferentes grupos de edad, debe utilizarse la aleatorización es-

tratificada. En esta situación, se crean estratos o grupos y la aleatorización se efectúa por separado para los sujetos en cada estrato. Cuando los sujetos son elegibles para el ensayo, se determina su estrato apropiado y reciben la asignación de números aleatorios dentro de este estrato. Por ejemplo, los pacientes pueden clasificarse según el sexo y la edad (menores de 50 y 50 o más), dando un total de cuatro estratos. Dentro de cada estrato, cada paciente se asignará aleatoriamente al grupo control o de intervención. Esto puede hacerse utilizando la aleatorización simple o por bloques permutados aleatoriamente.

La aleatorización estratificada presenta la ventaja de garantizar el equilibrio entre los grupos, como mínimo en relación a las características que determinan los estratos. La utilización de este método de aleatorización en el ejemplo descrito anteriormente asegurará que el grupo de intervención y control estén equilibrados respecto al sexo y la edad. Si no se hubiera empleado la estratificación, el investigador corría el riesgo de que el azar produjera un desequilibrio en lo que respecta a estos importantes factores, especialmente si el número de sujetos en el ensayo era pequeño. El inconveniente de la aleatorización estratificada es que es administrativamente difícil y complicada de ejecutar.

Diseño de pares apareados

Un diseño de pares apareados es un caso especial de aleatorización estratificada en la que los estratos son de tamaño 2. Los individuos (o comunidades) se agrupan en pares, lo más parecidos posibles en las variables de confusión, de tal modo que en ausencia de una intervención, se esperaría que presentaran un riesgo similar de la enfermedad en estudio. La intervención se asigna al azar a un miembro de cada par y el otro miembro actúa como control.

En ensayos grandes no es necesario agrupar, ya que es probable que cualquier desequilibrio entre los grupos de intervención, respecto a los factores de riesgo, en la aparición de los resultados de interés tienda a igualarse. Además, es posible ajustar cualquier desequilibrio residual durante el análisis del estudio sin pérdidas importantes de poder estadístico.

En ensayos pequeños puede producirse un desequilibrio más importante que puede ser difícil de ajustar en el análisis. Es un problema habitual en los ensayos en los que se distribuyen comunidades aleatoriamente, ya que en estos estudios no es habitual poder incluir un gran número de comunidades (más de 20) (Ejemplo 7.7). En estas situaciones es probable que el apareamiento de comunidades similares (es decir, comunidades en las que es probable que las tasas de la enfermedad sean similares en ausencia de las intervenciones aplicadas) antes de la asignación de las intervenciones sea una estrategia útil.

Ejemplo 7.7. En el estudio de intervención comunitario para dejar de fumar (COMMIT) mencionado anteriormente en el Ejemplo 7.3, dentro de cada par, las comunidades se aparearon por factores tales como las características demográficas de la población (p.e., tamaño de la población, edad, sexo y etnia), grado de urbanización, factores socioeconómicos, prevalencia del hábito de fumar y acceso a medios y servicios sanitarios. Las dos comunidades apareadas estaban lo suficientemente próximas geográficamente como para facilitar la monitorización a los investigadores, pero no tanto como para que las actividades educacionales en la comunidad de intervención afectaran a la comunidad control. Una comunidad de cada par se asignó aleatoriamente para recibir el programa de educación sanitaria y la otra a recibir sólo vigilancia epidemiológica (COMMIT Research Group, 1991).

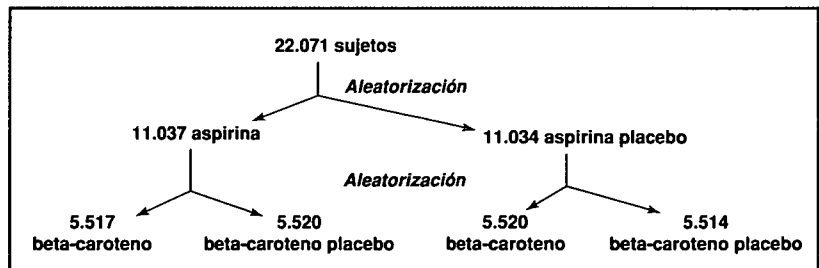
7.9.3 Algunos diseños experimentales especiales

Diseño factorial

Una técnica para mejorar la eficacia en los estudios de intervención es probar dos o más hipótesis simultáneamente en un diseño factorial. Un ensayo con dos hipótesis puede utilizar un diseño factorial de dos por dos, en el que los sujetos primero se distribuyen aleatoriamente a la intervención A o B para tratar la primera hipótesis, y después dentro de cada grupo de intervención se realiza otra distribución aleatoria a las intervenciones C y D para evaluar la segunda cuestión (Ejemplo 7.8).

Ejemplo 7.8. El Estudio sobre la Salud de los Médicos descrito en el Ejemplo 7.6 utilizó un diseño factorial 2×2 . Los médicos se asignaron a uno de los cuatro grupos como se muestra en la Figura 7.3:

Figura 7.3. Diagrama del diseño factorial del Estudio sobre la Salud de los Médicos.



Este diseño permitió el estudio de dos preguntas diferentes relacionadas con dos enfermedades diferentes: 1) ¿reduce la aspirina el riesgo de enfermedades cardiovasculares? 2) ¿reduce el beta-caroteno el riesgo de cáncer? Además, fue posible examinar el efecto combinado de los dos fármacos sobre la aparición de estas dos enfermedades (Hennekens et al., 1985; Stampfer et al., 1985).

La principal ventaja del diseño factorial es su capacidad de responder a dos o más preguntas en un único ensayo con sólo un pequeño aumento en el coste. Además, el uso de un diseño factorial permite también la evaluación de interacciones entre intervenciones, lo que no puede realizarse en un estudio de un único factor.

Ensayos cruzados

La mayoría de los ensayos tiene un diseño *paralelo*, es decir, un grupo de sujetos recibe la intervención y otro grupo paralelo recibe el tratamiento estándar o placebo. Por el contrario, en los *ensayos cruzados* cada sujeto actúa como su propio control recibiendo como mínimo dos intervenciones diferentes (p.e. un fármaco nuevo (tratamiento A) frente al fármaco estándar (tratamiento B)) a tiempos diferentes durante el ensayo (Figura 7.4). El orden en que cada individuo los recibe (A antes de B ó B antes de A) debe determinarse por asignación aleatoria.

Debe haber un periodo de 'lavado' ('wash-out') entre cada una de las intervenciones para evitar 'efectos de arrastre' (también denominados 'efectos *carry-over*' ó 'efectos *spillover*'), es decir, para asegurar que no se solapan los efectos de la primera y la segunda intervención. Por consiguiente, este diseño sólo es adecuado cuando ninguna de las intervenciones tiene efectos a largo plazo.

La principal ventaja de los ensayos cruzados es que cada sujeto se compara consigo mismo y, por tanto, la confusión se elimina de la comparación de los efectos de los dos tratamientos (siempre que no exista 'efecto de arrastre'). Este diseño también aumenta la precisión estadística en la comparación porque elimina la variabilidad inter-sujeto en la respuesta del resultado. Por tanto, son necesarios menos sujetos que en el correspondiente ensayo paralelo.

Los ensayos cruzados se utilizan principalmente en las fases iniciales de evaluación de nuevos fármacos en las que se investigan sus propiedades farmacocinéticas en voluntarios sanos. No son apropiados para evaluar los

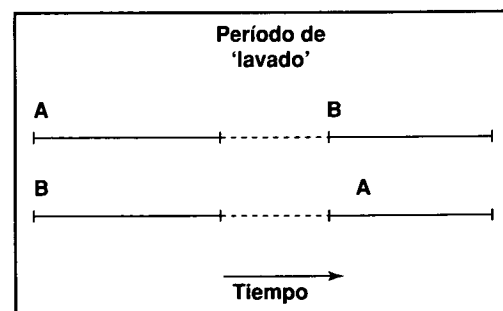


Figura 7.4. Diagrama de un ensayo cruzado.

Ejemplo 7.9. Treinta y un pacientes con un diagnóstico de tumor metastásico de células germinales, que recibieron un tratamiento durante 4 días de quimioterapia con cisplatino, se incluyeron en un ensayo cruzado, aleatorizado a doble ciego. El objetivo del ensayo fue evaluar si el ondansetron oral (un antagonista de la serotonina) con dexametasona era más eficaz que el ondansetron oral con placebo en el control de la emesis asociada a la quimioterapia. Durante el primer tratamiento de quimioterapia, los pacientes se distribuyeron aleatoriamente a uno de los grupos del ensayo. A los participantes se les administró uno de los tratamientos durante ocho días. Un segundo tratamiento de quimioterapia se administró 14 días después del inicio del primero durante el cual los pacientes se cruzaron al régimen antiemético alternativo (Smith et al., 1991).

efectos a largo plazo de un tratamiento, ya que la naturaleza del diseño implica que debe limitarse el periodo de tratamiento.

7.9.4 Conclusiones

Cabe recordar que la asignación de los sujetos a los grupos de estudio sólo debe realizarse después de haber determinado que el individuo es elegible y que está dispuesto a participar. De otro modo, los sujetos que se niegan a participar o que se retiran del estudio (porque el tratamiento no es adecuado, etc.) tendrán que excluirse después de la aleatorización y, como resultado, puede que los grupos ya no sean comparables. Esto se ilustra en el Ejemplo 7.10.

Ejemplo 7.10. *El efecto del cribaje del cáncer de mama sobre la mortalidad se examinó en un ensayo aleatorizado. Mujeres de 40-64 años miembros de Plan de Seguridad Sanitaria de Nueva York se distribuyeron aleatoriamente en dos grupos: un grupo de intervención, al que se ofrecieron pruebas de cribaje (exámenes clínicos y mamografía) a intervalos anuales; y un grupo control, que siguió recibiendo la atención médica habitual. Había aproximadamente 31.000 mujeres en cada grupo. Los grupos eran muy parecidos respecto a un gran número de características, (Shapiro, 1977). La Tabla 7.2 muestra la mortalidad por causas diferentes al cáncer de mama para cada uno de los dos grupos de estudio.*

Tabla 7.2. Mortalidad por causas diferentes al cáncer de mama al final de los primeros 5 años de seguimiento^a.

Grupo de estudio	Nº de muertes	Tasa de mortalidad (por 10.000 p-as)
Grupo de intervención	850	57
Cribadas	421	42
Rehusaron	429	86
Grupo control	877	58

^a Datos de Shapiro (1977).

La Tabla 7.2 muestra que había una gran diferencia en la mortalidad por causas diferentes al cáncer de mama entre las mujeres que recibían la intervención (es decir, que se sometieron a cribaje) y las del grupo control (42 frente a 58 por 10.000 personas-años). Debido a que la intervención en estudio (es decir, cribaje de mama) no debería haber afectado a la mortalidad por causas diferentes al cáncer de mama, la diferencia observada parece indicar que los dos grupos eran diferentes con relación a características basales importantes. No obstante, si se incluyen en el grupo de intervención aquellos que se negaron a participar después de la aleatorización, ya no se observan diferencias de mortalidad entre los dos grupos (57 frente a 58 por 10.000 personas-años).

La asignación aleatoria no garantiza necesariamente que los grupos sean similares. Pueden aparecer discrepancias al azar entre los grupos, especialmente si el número de unidades a asignar (p.e. individuos, familias, comu-

nidades) es relativamente bajo. Por tanto, es esencial obtener datos basales de los sujetos. Estos datos basales deben incluir todas las variables que se sabe o se cree que afectarán al resultado de interés y pueden utilizarse para comprobar el grado de similitud de los grupos. Si los grupos de estudio son diferentes, existen técnicas estadísticas que pueden tratar este problema proporcionando resultados 'ajustados' por las diferencias basales entre los grupos de estudio (Capítulos 13-14). Esto obviamente es imposible si no se han recopilado los datos basales.

7.10 Monitorización del cumplimiento y de las reacciones secundarias

El problema de conseguir y mantener un alto cumplimiento es un aspecto importante en el diseño y realización de un ensayo. Esto es debido a que el no-cumplimiento hace que una fracción del grupo de tratamiento no lo reciba y actúe de hecho como el grupo control. Entonces, los grupos de intervención y de comparación son más parecidos y, por consiguiente, se reduce la capacidad de detectar diferencias reales entre los grupos. Un cierto grado de no-cumplimiento es aceptable en ensayos de tipo pragmático ya que están destinados a estimar la eficacia de la intervención en condiciones reales. Por el contrario, el objetivo de los ensayos científicos es estimar el máximo beneficio potencial de la intervención en circunstancias ideales, incluyendo un cumplimiento del 100%. Una manera de aumentar el cumplimiento es utilizar una 'fase de inclusión' antes de la distribución aleatoria como se ilustró en el Ejemplo 7.6.

Los niveles de cumplimiento deben determinarse y monitorizarse a lo largo del estudio. Esto puede realizarse utilizando auto-informes. Este método presenta el inconveniente de que confía exclusivamente en la opinión y en la memoria del sujeto. La devolución de la medicación no utilizada (p.e. comprimidos) al investigador a intervalos regulares se ha utilizado en ensayos en que se administran fármacos o sustancias activas. Este método supone que el sujeto ha tomado toda la medicación que no ha devuelto. Los auto-informes a veces se validan con las medidas de laboratorio. Aparte de ser caros, estos métodos también tienen un valor limitado ya que normalmente miden el cumplimiento actual y no a largo plazo.

En el Ejemplo 7.11, las medidas de alfa-tocoferol y beta-caroteno realizadas tres años después de la inclusión en el ensayo fueron mucho más altas que el valor basal en aquellos sujetos que recibieron estos principios activos pero cambiaron poco en los que no los recibieron (Tabla 7.3). Estas conclusiones indican unos niveles altos de cumplimiento.

En la mayoría de los ensayos, una proporción de los participantes inevitablemente interrumpirá el tratamiento por una razón u otra (olvidarse de tomar los fármacos, desarrollo de efectos secundarios, etc.). En términos generales es preferible un menor nivel de cumplimiento a un total no-cumplimiento. Más aun, como se comentará más adelante en el Apartado 7.12, en todo estudio de intervención cada sujeto aleatorizado debe incluirse en

Ejemplo 7.11. El estudio de prevención del cáncer con alfa-tocoferol y beta-caroteno [The Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study] fue un ensayo aleatorizado, doble ciego, controlado por placebo para determinar si el suplemento diario con alfa-tocoferol, beta-caroteno o ambos reduciría la incidencia de cáncer de pulmón y otros cánceres. Un total de 29.133 fumadores varones de 50 a 69 años se asignaron aleatoriamente a uno de los cuatro regímenes: alfa-tocoferol (50 mg al día) solo; beta-caroteno (20 mg al día) solo; alfa-tocoferol y beta-caroteno; o placebo. El seguimiento duró de cinco a ocho años. El cumplimiento se evaluó por recuento de las cápsulas restantes en cada visita, determinación de los niveles séricos de alfa-tocoferol y beta-caroteno después de tres años de administración del suplemento (Tabla 7.3), y por determinación en muestras de suero aleatorias durante el estudio (The Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study Group, 1994).

Tabla 7.3.

Concentraciones séricas (miligramos por mililitro) de alfa-tocoferol y beta-caroteno al inicio del estudio y a los tres años de administración de suplemento, según el grupo de estudio^a.

Tratamiento recibido	Nº de sujetos	Media	Percentil 20	Percentil 80
<i>Niveles séricos basales de alfa-tocoferol</i>				
Alfa-tocoferol	14.472	11,5	9,3	14,2
No alfa-tocoferol	14.469	11,4	9,3	14,1
<i>Niveles séricos de alfa-tocoferol a los tres años</i>				
Alfa-tocoferol	11.332	17,3	14,3	21,1
No alfa-tocoferol	11.258	12,4	10,2	15,1
<i>Niveles séricos basales de beta-caroteno</i>				
Beta-caroteno	14.460	0,17	0,10	0,29
No beta-caroteno	14.460	0,17	0,10	0,29
<i>Niveles séricos de beta-caroteno a los tres años</i>				
Beta-caroteno	11.276	3,0	1,6	4,5
No beta-caroteno	11.314	0,18	0,10	0,30

^a Datos de ATBC Cancer Prevention Study Group (1994).

el análisis preliminar, pues es esencial obtener información del seguimiento lo más completa posible de los que se han retirado del programa de intervención. Los investigadores deben seguir a estos individuos a lo largo del ensayo y obtener información sobre los resultados relevantes de la misma manera que para los sujetos que siguen cumpliendo.

A veces los asignados a un grupo pueden elegir pertenecer al grupo de intervención por iniciativa propia. Por ejemplo, aquellos asignados al grupo control pueden adoptar el tratamiento activo. Es importante reducir esta «contaminación» en la medida que sea posible. Una manera es diseñar el ensayo de modo que se reduzcan las oportunidades de contaminación. En el Ejemplo 7.4 se eligieron vuelos en lugar de individuos como unidad de aleatorización para reducir la contaminación. Asimismo, en un ensayo comunitario para evaluar el impacto de un programa antitabaco es importan-

te que las comunidades de intervención y control sean unidades geográficamente diferenciadas con poblaciones estables y sin migración entre ellas (como en el Ejemplo 7.7). Pero algunas veces es inevitable un cierto grado de contaminación debido a razones que están completamente fuera de control de los investigadores.

Ejemplo 7.12. *En Huixian, República Popular de China, se realizó un estudio de intervención aleatorizado a doble ciego, para determinar si el tratamiento combinado con retinol, riboflavina y zinc durante un año podría reducir la prevalencia de lesiones precancerosas de esófago. Un total de 610 sujetos de 35-64 años se distribuyeron aleatoriamente al tratamiento activo o al placebo. El cumplimiento fue muy bueno. Al final del ensayo, la prevalencia de las lesiones de esófago fue similar en los dos grupos: 45,3% en el grupo placebo y 48,9% en el grupo de intervención (Muñoz et al., 1985).*

Al final del ensayo del Ejemplo 7.12 no se observaron diferencias en la prevalencia de las lesiones de esófago entre los grupos placebo y tratado. Las mediciones de laboratorio efectuadas en el momento de la inclusión en el estudio y dos meses después no sólo confirmaron que había un aumento en los niveles de vitaminas en el grupo de tratamiento activo, sino que también mostraron que los niveles de retinol habían mejorado en el grupo placebo. El cambio en el grupo placebo probablemente fue debido a un mejor acceso a frutas y verduras frescas (Muñoz *et al.*, 1985).

Un problema similar de contaminación entre grupos se produjo en el ensayo del programa antitabaco descrito en el Ejemplo 7.2. Durante el periodo de 20 años que duró el estudio, hubo una disminución progresiva en la prevalencia de fumadores del grupo control (Rose y Colwell, 1992) reflejando un aumento en la sensibilización sobre las consecuencias negativas de fumar en la salud. Esta contaminación hizo que los dos grupos fueran más parecidos y redujo la capacidad del estudio para medir los efectos de dejar de fumar en la salud.

También es importante controlar las reacciones secundarias adversas que puedan desarrollarse. Debe establecerse un mecanismo de control que permita la ruptura del código de aleatorización si algún sujeto desarrolla reacciones adversas graves. La monitorización de las reacciones adversas no sólo es necesaria para garantizar la seguridad de los participantes del estudio sino que ayudará además a evaluar los beneficios y riesgos reales de la intervención en estudio.

7.11 Evaluación de los resultados

Los resultados de interés deben definirse claramente antes del inicio del ensayo. La elección del resultado de interés tiene implicaciones importantes para la duración del ensayo. La mayoría de los ensayos de campo en epidemiología del cáncer están destinados a reducir el riesgo de esta enfermedad. Debido a que los ensayos de campo se realizan con personas exentas

de la enfermedad, la probabilidad de desarrollar cáncer es relativamente baja y puede no observarse durante varias décadas. El problema no es tan acentuado en ensayos clínicos de cáncer ya que la mayoría de los resultados de interés (p.e. recidiva o muerte) tienden a producirse con una alta frecuencia.

Una manera de acortar la duración del ensayo es seleccionar una población con un mayor riesgo de desarrollar el resultado de interés. Por ejemplo, en un ensayo de campo que evalúa el impacto de un programa antitabaco sobre las tasas de mortalidad por cáncer de pulmón tiene sentido excluir a todas las personas menores de 45 años ya que el cáncer de pulmón aparece raramente en esas edades.

Otra posibilidad es utilizar variables intermedias como sustitutos del cáncer, es decir, utilizar como resultado un acontecimiento biológico que se encuentra en la vía causal entre la exposición y el cáncer. Los estudios que utilizan variables intermedias son más rápidos, más pequeños y menos caros que los estudios que utilizan la enfermedad clínica como resultado. No obstante, la relevancia de los resultados del ensayo respecto al cáncer dependerá de la fuerza de la asociación entre la variable intermedia y el cáncer.

Ejemplo 7.13. *Se inició recientemente un ensayo multicéntrico, aleatorizado para probar un régimen dietético en la disminución del riesgo de recurrencia de pólipos del intestino grueso. Los pacientes con uno o más pólipos adenomatosos confirmados por histología, a los que se les habían eliminado los pólipos por colonoscopia, se asignaron aleatoriamente a recibir una dieta habitual o a recibir educación y consejo nutricional destinados a un cambio del estilo de vida hacia una dieta pobre en grasas, rica en fibra y enriquecida con frutas y verduras. El ensayo incluyó 2.000 pacientes con un seguimiento de 4 años; la colonoscopia debe repetirse al cabo de 1 y 4 años. El resultado principal es la recidiva de adenomas. El ensayo se basa en el postulado de que la mayoría de los cánceres de intestino grueso se producen a partir de pólipos adenomatosos. Los resultados del ensayo deben proporcionar evidencias sobre la capacidad de que el cambio de la dieta pueda afectar la recidiva de pólipos adenomatosos y, por lo tanto de afectar indirectamente la incidencia de cáncer de intestino grueso (Chlebowski y Grosvenor, 1994).*

En el Ejemplo 7.13 sólo son necesarios 2.000 individuos y cuatro años de seguimiento, un número considerablemente menor del que habría sido necesario para un ensayo con cáncer de intestino grueso como resultado final. La suposición fundamental en estos estudios es que la relación observada entre la exposición (p.e. la dieta) y la variable intermedia (p.e. los pólipos) refleja una relación similar entre la exposición y el cáncer *per se*. Está claro que esta suposición se tiene que validar antes de aceptar una variable intermedia como sustituto del cáncer (Lippman *et al.*, 1990).

Los resultados deben determinarse de tal modo que el sesgo de medición se reduzca al mínimo posible. Las técnicas a ciegas o de enmascaramiento ayudan a conseguirlo. Cuando no existe una intervención o tratamiento estándar para utilizar en un estudio a ciegas en comparación con la nueva intervención, deben utilizarse placebos para mantenerlo 'ciego'. El placebo debe ser lo más parecido posible a la propia intervención (respecto al aspecto, sabor, etc.). Siempre que sea posible, ni el paciente ni el investigador deben saber quién está asignado a cada grupo hasta el final del ensayo. Estos diseños «doble ciego» (tanto el investigador como los participantes están «cegados») eliminan la posibilidad de que la información sobre el tipo de intervención asignada a un individuo pueda afectar la manera de tratar o controlar al individuo durante el ensayo, la manera de que un individuo responda a la intervención o la evaluación de un individuo al final del ensayo. Un ensayo doble ciego puede ser irrealizable si se trata de evaluar programas que impliquen cambios sustanciales en el estilo de vida, tales como hacer ejercicio, fumar cigarrillos, o cambiar la dieta, intervenciones quirúrgicas o fármacos con reacciones adversas características. En estas circunstancias, la única posibilidad es un estudio a «*simple ciego*» (el investigador sabe a qué grupo pertenece un participante pero el participante no, o viceversa).

Cuanto más subjetivo sea el resultado en estudio, mayor será la justificación para realizar un ensayo a doble ciego. Por ejemplo, si tratamos con respuestas extremadamente subjetivas, tales como el alivio del dolor o la mejoría del estado psicológico, el uso del doble ciego es crucial para la validez de las medidas del resultado. Cuando el resultado de un ensayo es más objetivo, por ejemplo, vida o muerte, o quizás el nivel de algunas sustancias en sangre u orina, la necesidad de un ensayo doble ciego, obviamente, es menos importante.

La fuerza principal de un diseño doble ciego es eliminar el potencial para el sesgo de medida. Por supuesto, una limitación concomitante es que estos ensayos normalmente son más complejos y difíciles de realizar. Deben establecerse procedimientos para romper de manera inmediata el «cegado» del médico en el caso de que algún participante presente reacciones adversas graves u otras emergencias clínicas en las que esta información parece esencial.

7.12 Análisis

7.12.1 Tipos de análisis

Existen dos métodos principales para el análisis de un ensayo en relación a quién debe incluirse o no en el ensayo. El análisis de '*intención de tratar*' se basa en los desenlaces que se producen durante todo el periodo de seguimiento y en que los sujetos asignados originalmente a cada grupo permanezcan o no en el grupo de intervención. La alternativa es el análisis '*en tratamiento aleatorizado*', que se limita a los desenlaces observados mientras los sujetos están en su tratamiento asignado. La exclusión del análisis de sujetos aleatorizados puede dar lugar a un sesgo importante que puede surgir de

los diferentes grados de participación en los grupos de intervención y control, y del hecho que los individuos retirados o perdidos durante el seguimiento son normalmente diferentes de los que participan hasta el final. El 'análisis de intención de tratar' es la forma correcta de analizar los datos, incluyendo en la comparación de los resultados a todos los sujetos asignados originalmente a cada grupo (incluso aquellos que no recibieron o que interrumpieron la intervención especificada). No obstante, si esta aproximación se toma de forma estricta puede llegar a diluir el efecto real de la intervención.

7.12.2 Análisis seriados

El análisis de los resultados de un ensayo a medida que se van acumulando datos es una forma importante de monitorizar su progreso. Los análisis administrativos de los números de participantes incluidos cada día o semana y de los datos recogidos por los distintos trabajadores de campo son importantes para el control de calidad del estudio.

En estudios grandes, con frecuencia se establece un comité independiente de monitorización de datos para mantener el código de aleatorización del estudio y monitorizar los resultados del ensayo a medida que se obtienen o a intervalos fijos durante el estudio. Este comité tiene poder para interrumpir la inclusión de nuevos participantes si existen evidencias de un riesgo sustancial de reacciones adversas asociadas con alguna de las intervenciones en estudio. Asimismo, si se acumulan evidencias de que una intervención es considerablemente mejor que las otras (o una considerablemente peor), el comité puede recomendar que se interrumpa la fase de intervención del ensayo y que se administre a todos los participantes el tratamiento o intervención mejor (o menos dañina). Sería muy difícil para los investigadores permanecer objetivos e imparciales si tuvieran que tomar ellos estas decisiones.

Ejemplo 7.14. *El Estudio de Eficacia del Beta-Caroteno y Retinol [The Beta-Carotene and Retinol Efficacy Care (CARET)] es un ensayo multicéntrico, aleatorizado, a doble ciego, controlado por placebo, iniciado en 1983 para evaluar si una combinación de beta-caroteno y retinol (vitamina A) podía reducir la incidencia de cáncer de pulmón en poblaciones de alto riesgo. Un total de 14.254 fumadores y 4.060 trabajadores expuestos a amianto se distribuyeron aleatoriamente a la intervención (beta-caroteno y vitamina A) o placebo. El diseño del ensayo estipuló que la administración de las intervenciones debía durar hasta finales de 1997. No obstante, en enero de 1996, el comité de monitorización de datos decidió terminar la intervención porque se observó que había un aumento del 28% (intervalo de confianza del 95%, 0,4% a 57%; $P = 0,02$) en el riesgo de cáncer de pulmón en el grupo de intervención en comparación con el grupo placebo. Se espera que el seguimiento continúe durante otros cinco años para identificar posibles casos adicionales (Omenn et al., 1996).*

En el Ejemplo 7.14 la fase de intervención del ensayo se terminó antes. El seguimiento post-intervención de los sujetos del estudio continuará durante cinco años para identificar casos adicionales de cáncer de pulmón y evaluar los efectos a largo plazo de la intervención.

7.12.3 Análisis final

El primer paso en el análisis de un ensayo es examinar las características de los dos (o más) grupos al inicio del estudio y evaluar si son comparables, verificando si la distribución aleatoria dio lugar a la formación de grupos comparables y equilibrados (Ejemplo 7.15). Esta comparación debe constituir la primera tabla del apartado de resultados de un artículo.

Ejemplo 7.15. Las características basales de los participantes en el ensayo CARET descrito en el Ejemplo 7.14 se presentan en la Tabla 7.4. El grupo de intervención y el grupo placebo eran similares respecto a un número considerable de factores que podían haber influenciado al principal resultado del estudio, es decir, la incidencia de cáncer de pulmón.

Característica	Trabajadores expuestos al amianto		Grandes fumadores	
	Intervención	Placebo	Intervención	Placebo
Nº aleatorizados	2.044	2.016	7.376	6.878
Edad (años)^b	57±7	57±7	58±5	58±5
Mujeres^c	0	0	3.208 (43)	3.081 (45)
Raza o grupo étnico^c				
Blanca	1.805 (88)	1.775 (88)	7.000 (95)	6.487 (94)
Negra	152 (7)	153 (8)	103 (1)	122 (2)
Hispana	36 (2)	43 (2)	101 (1)	95 (1)
Otros/desconocido	51 (2)	45 (2)	172 (2)	174 (3)
Situación de hábito tabáquico^c				
Nunca fumador	68 (3)	64 (3)	0	0
Ex-fumador	1.195 (58)	1.175 (58)	2.473 (34)	2.331 (34)
Fumador actual	781 (38)	777 (39)	4.903 (66)	4.547 (66)
Cigarrillos fumados/día^b				
Ex-fumadores	25±12	25±12	28±11	28±11
Fumadores act.	24±10	25±10	24±9	24±8
Paquetes-años fumando^b	43±24	42±24	50±21	49±20
(sólo para ex-fumadores y fumadores actuales)				
Años desde que se dejó de fumar^b	10±8	10±8	3±2	3±2
(sólo para ex-fumadores)				

^a Datos de Omenn *et al.* (1996)

^b Media ± desviación estándar.

^c Número (%).

Tabla 7.4.

Distribución de las características basales de grandes fumadores y trabajadores expuestos al amianto que participaron en el ensayo CARET, según el tipo de intervención recibida^a.

Con frecuencia se realizan ensayos estadísticos para evaluar si las diferencias basales entre los grupos en estudio son importantes. Una prueba estadística proporciona la probabilidad de encontrar una diferencia como mínimo tan grande como la observada debido únicamente al azar. No obstante, se sabe que todas las diferencias observadas, independientemente de su magnitud, se han producido sólo por azar ya que los sujetos estaban aleatorizados. Por lo tanto; *las pruebas estadísticas son superfluas e inadecuadas para evaluar si los grupos de estudio tienen características basales similares.*

El examen de las características basales de los grupos puede ayudar también a desvelar problemas desconocidos que puedan haberse producido durante el proceso de aleatorización. Por ejemplo, si las características basales de los grupos son muy diferentes, debe comprobarse todo el proceso de aleatorización para excluir la posibilidad de que pueda haber sido mal aplicado el proceso de inclusión de los sujetos en el ensayo.

Tras evaluar la comparabilidad de los grupos en estudio, el investigador debe determinar si la intervención sirvió para algo. Los dos grupos se comparan y se evalúa el tamaño de las diferencias. En general, los principales resultados de un ensayo pueden presentarse en una tabla similar a la mostrada en la Tabla 7.5.

Si todos o prácticamente todos los participantes se incluyen y se retiran del ensayo a la vez, entonces puede calcularse el riesgo (Tabla 7.5a). Por ejemplo, si el periodo de seguimiento es uniforme de 3 años, entonces puede calcularse el riesgo a los 3 años para cada grupo de estudio. Los dos grupos de estudio pueden compararse calculando las razones y las diferencias de riesgo como medidas del riesgo relativo y del exceso de riesgo, respectivamente.

En el Ejemplo 7.16, prácticamente todos los participantes se incluyeron en el ensayo en el mismo momento del tiempo y se siguieron hasta el final del estudio. Si se asume que, en promedio, las muertes en cada grupo de tratamiento se produjeron en momentos similares en el tiempo, el cálculo del riesgo como medida de aparición de la enfermedad es adecuado. Los resultados de este ensayo fueron compatibles con la hipótesis nula de no diferencias entre tratamientos en el riesgo de cáncer de pulmón, es decir, una razón de riesgos igual a 1 (o una diferencia de riesgo igual a cero).

Sin embargo, muchos estudios de intervención implican periodos variables de seguimiento. La inclusión en el ensayo puede llevar varios años y si el seguimiento se termina en una fecha específica, los participantes se habrán observado durante intervalos de tiempo diferentes. Además, algunos sujetos se pierden durante el seguimiento o mueren en diferentes intervalos de tiempo y en consecuencia habrán formado parte del ensayo durante diferentes periodos de tiempo.

El cálculo de persona-tiempo de observación como denominador para el cálculo de las tasas es el método utilizado habitualmente en estudios de intervención cuando deben tenerse en cuenta periodos de observación variables (que son el resultado de la entrada y el abandono de personas en el es-

(a)	Exposición		
		Intervención	Control
Resultado	Sí	<i>a</i>	<i>b</i>
	No	<i>c</i>	<i>d</i>
Riesgo en el grupo de intervención $(\rho_1) = a/(a+c)$			
Riesgo en el grupo control $(\rho_0) = b/(b+d)$			
Razón de riesgos = ρ_1/ρ_0			
Diferencia de riesgos ^a = $\rho_0 - \rho_1$			
(b)	Exposición		
		Intervención	Control
Nº de casos	<i>a</i>	<i>b</i>	
Persona-tiempo a riesgo	<i>y</i> ₁	<i>y</i> ₀	
Tasa en el grupo intervención $(r_1) = a/y_1$			
Tasa en el grupo control $(r_0) = b/y_0$			
Razón de tasas = r_1/r_0			
Diferencia de tasas ^a = $r_0 - r_1$			
^a En la mayoría de los ensayos la intervención es protectora y por tanto debe calcularse la diferencia de riesgo y de tasas como se indica (ver Apartado 5.2.2.). Si la intervención está asociada con un aumento de la incidencia, la diferencia de riesgo y de tasas deben calcularse como $\rho_0 - \rho_1$ o $r_0 - r_1$, respectivamente			

Tabla 7.5.
Análisis de un estudio de intervención:
(a) por riesgos; (b) por tasas.

Ejemplo 7.16. En el Estudio sobre la Salud de los Médicos descrito en los Ejemplos 7.6 y 7.8, 22.071 médicos varones de los EE.U.U. de edades comprendidas entre 40 a 84 años se distribuyeron aleatoriamente para recibir uno de entre cuatro tratamientos: 1) aspirina más beta-caroteno placebo; 2) beta-caroteno más aspirina placebo; 3) los dos principios activos; 4) los dos placebos. El comité externo de monitorización de datos interrumpió prematuramente, en 1988, la aleatorización del componente con aspirina del ensayo porque se observó una reducción del 44% ($P < 0,001$) en el riesgo de un primer infarto de miocardio en los sujetos que tomaban aspirina. El componente aleatorizado de beta-caroteno continuó ininterrumpidamente hasta el final, 31 de diciembre de 1995. Un total de 11.036 médicos recibieron beta-caroteno y 11.035 recibieron beta-caroteno placebo y menos de un 1 por ciento se perdieron en el seguimiento. Uno de los principales objetivos de este componente del estudio era evaluar si el beta-caroteno reduce la incidencia de cáncer de pulmón (Hennekens et al., 1996). La Tabla 7.6 muestra los resultados.

		Grupo de estudio	
		Beta-caroteno	Placebo
Cáncer de pulmón	Sí	82	88
	No	10.954	10.947
	Total	11.036	11.035

^a Datos de Hennekens *et al.*, 1996

$$p_1 = 82/11.036 = 0,00743 = 7,43 \text{ por } 1.000$$

$$p_0 = 88/11.035 = 0,00797 = 7,97 \text{ por } 1.000$$

$$\chi^2 = 0,21; 1 \text{ g.l.}; P > 0,50$$

$$\text{Razón de riesgos } (p_1/p_0) = 0,93$$

Intervalo de confianza al 95% para la razón de riesgos = 0,69 a 1,26

Diferencia de riesgo ($p_0 - p_1$) = 7,97 por 1.000 - 7,43 por 1.000 = 0,54 por 1.000

Intervalo de confianza al 95% para la diferencia de riesgo = - 1,77 por 1.000 a 2,85 por 1.000

(Las pruebas estadísticas y los intervalos de confianza se calcularon utilizando las fórmulas presentadas en el Anexo 6.1.).

Tabla 7.6.

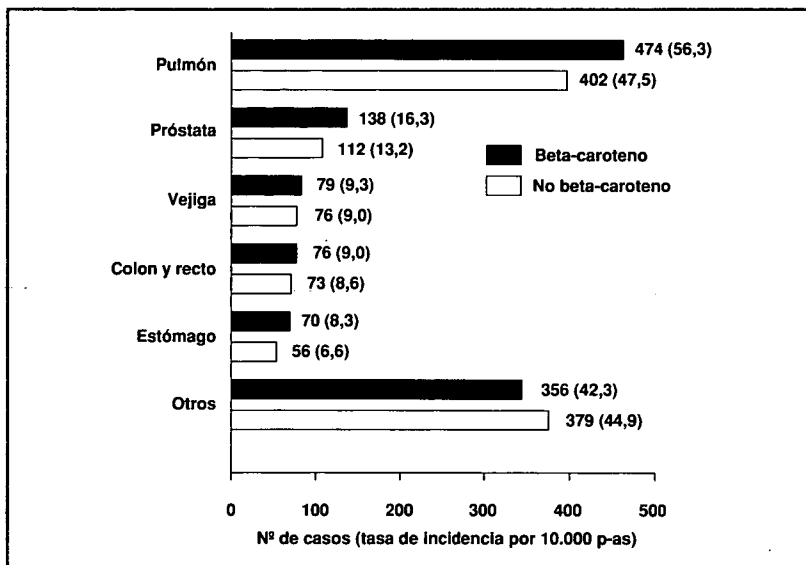
Distribución de la incidencia de casos de cáncer de pulmón en el Estudio sobre la Salud de los Médicos, según el grupo de tratamiento^a.

tudio a diferentes edades y tiempos). Los resultados del ensayo pueden presentarse tal como se muestra en la Tabla 7.5.b, y las razones y diferencias de tasas pueden calcularse como medidas del efecto relativo y absoluto, respectivamente.

En el Ejemplo 7.17, la duración del seguimiento varió entre sujetos. Por tanto es más adecuado calcular las unidades de persona-tiempo a riesgo y tasas como medida de aparición de la enfermedad. Los resultados de este estudio no justificaron la hipótesis de que el beta-caroteno reduce la incidencia de cáncer de pulmón. De hecho, proporcionan evidencia de que la administración de beta-caroteno puede aumentar el riesgo de cáncer de pulmón.

Ejemplo 7.17. El objetivo del ensayo ATBC (descrito en el Ejemplo 7.11) fue evaluar si el suplemento diario con alfa-tocoferol, beta-caroteno o los dos reducía la incidencia de cáncer de pulmón y otros cánceres. Un total de 29.133 fumadores varones de 50 a 69 años del sudoeste de Finlandia se incluyeron entre 1985 y 1988. El seguimiento de éstos continuó durante un período de 5 a 8 años (mediana = 6,1) hasta la muerte o el 30 de abril de 1993, cuando terminaba el ensayo (The ATBC Cancer Prevention Study Group, 1994). Los resultados por tipo de tratamiento recibido se muestran en la Figura 7.5.

Figura 7.5. Número de casos de cáncer y tasas de incidencia según localización y tipo de tratamiento recibido (reproducido con la autorización del Alpha-Tocoferol, Beta-Carotene Cancer Prevention Study Group, 1994).



Los resultados de la incidencia de cáncer de pulmón en función del suplemento de beta-caroteno pueden presentarse como se muestra en la Tabla 7.7.

Tabla 7.7. Distribución de la incidencia de casos de cáncer de pulmón y personas-años a riesgo según el tratamiento recibido^a.

	Grupo de estudio	
	Beta-caroteno	No beta-caroteno
Casos de cáncer de pulmón	474	402
Personas-años a riesgo	84.192	84.632

^a Datos del ATBC Cancer Prevention Study Group, 1994

$r_1 = 474/84.192 = 56,3$ por 10.000 p-as
 $r_0 = 402/84.632 = 47,5$ por 10.000 p-as
 $\chi^2 = 6,30$; 1 g.l.; $P = 0,01$
Razón de tasas (r_1/r_0) = 1,19
Intervalo de confianza al 95% para la razón de tasas = 1,04 a 1,36
Diferencia de tasas ($r_1 - r_0$) = 8,8 por 10.000 p-as
Intervalo de confianza al 95% para la diferencia de tasas = 1,9 a 15,7 por 10.000 p-as
(Las pruebas estadísticas y los intervalos de confianza se calcularon utilizando las fórmulas presentadas en el Anexo 6.1).

Las pruebas estadísticas y, principalmente, los intervalos de confianza para las medidas del efecto relativo y absoluto siempre deben calcularse y presentarse. En los Ejemplos 7.16 y 7.17 utilizamos las fórmulas presentadas en el Anexo 6.1.

La *fracción evitada* (véase Apartado 5.2.2) es otra medida importante en los estudios de intervención. Mide la proporción de casos de enfermedad que se evitaron gracias a esta intervención.

$$\text{Fracción evitada (\%)} = 100 \times [\text{diferencia de tasa (o riesgo) / tasa (o riesgo) en los no expuestos}]$$

Si el objetivo del ensayo es evaluar el valor de una vacuna, esta medida se denomina *eficacia de la vacuna*. Por ejemplo, si el riesgo de desarrollar una enfermedad en particular entre los vacunados era 40 por 10.000 y de 70 por 10.000 entre los no vacunados, la eficacia de la vacuna sería

$$\text{Eficacia de la vacuna (\%)} = 100 \times [(70 \text{ por } 10.000 - 40 \text{ por } 10.000) / 70 \text{ por } 100.000] = 43\%$$

Así, un 43% de los casos se habrían evitado entre los no vacunados si se hubieran vacunado.

Si en el análisis tienen que considerarse las diferencias basales entre los grupos en estudio, entonces debe utilizarse una de las técnicas estadísticas comentadas en el Capítulo 14.

Para evaluar la distribución del *tiempo hasta la aparición del suceso de interés* (p.e. tiempo desde el tratamiento hasta la muerte o tiempo desde el tratamiento hasta la recidiva), como ocurre en muchos ensayos clínicos, el método más adecuado es el *análisis de la supervivencia*. Las técnicas utilizadas en el análisis de la supervivencia proceden de los métodos de tablas de vida que se comentarán en el Capítulo 12.

Análisis de subgrupos

Es habitual entre los investigadores realizar análisis de subgrupos para evaluar si la intervención tiene algún efecto específico sobre subgrupos de individuos con determinadas características (p.e. varones, ancianos, pacientes con unos aspectos clínicos especiales, etc.). No obstante, estos análisis de subgrupos implican problemas importantes. Si los *subgrupos se definen según las características basales* de los pacientes, entonces el principal problema es la pérdida de capacidad del ensayo para detectar un efecto (es decir, pérdida de poder o precisión estadística (véase Capítulo 15)) ya que se basarán solamente en una pequeña proporción del número total de sujetos distribuidos aleatoriamente. Por otro lado, si se realizan análisis múltiples, puede preverse que algunos alcanzarán inevitablemente la 'significación estadística' simplemente al azar. Su interpretación dependerá en gran parte de la existencia de hipótesis *a priori* basadas en la plausibilidad biológica, en la existencia de evidencias de ensayos de laboratorio y en otras observaciones epidemiológicas.

Los análisis realizados en *subgrupos definidos basándose en las características individuales que se desarrollan tras la aleatorización* son mucho más preocupantes debido a que las variables de confusión potenciales ya no se distribuirán al azar entre los subgrupos. Por ejemplo, el análisis limitado a personas que alcanzaron una concentración sérica determinada de tratamiento activo o que desarrollaron una reacción adversa conocida (p.e. coloración amarilla de la piel al ingerir beta-caroteno) deben tratarse con precaución extrema. Sus observaciones nunca deben notificarse como resultados principales del ensayo sino sólo como observaciones interesantes que merecen ser investigadas en estudios diseñados específicamente para esto.

7.13 Interpretación

La interpretación de los resultados de un estudio de intervención bien realizado debe ser relativamente directa ya que los dos principales problemas de los estudios observacionales, el sesgo y la confusión, se reducen en gran parte utilizando un diseño experimental.

Esto no significa que los ensayos no tengan problemas. La historia del beta-caroteno y el cáncer de pulmón es un buen ejemplo de ello. Los datos de tres ensayos de quimioprofilaxis a gran escala realizados en países occidentales para evaluar esta cuestión ya han sido publicados. El estudio ATBC de prevención del cáncer (Ejemplo 7.11) se realizó en Finlandia para probar la hipótesis de que una ingesta elevada de beta-caroteno y alfa-tocoferol reducía el riesgo de cáncer de pulmón. Ésta era una hipótesis razonable dada la evidencia acumulada por estudios epidemiológicos observacionales que indicaban que el consumo de alimentos ricos en beta-caroteno estaba asociado a un menor riesgo de cáncer de pulmón. Los resultados de este ensayo no proporcionaron evidencia de que el beta-caroteno (o alfa-tocoferol) fuera beneficioso en la prevención de esta enfermedad. De hecho, los hombres que tomaron beta-caroteno presentaban un aumento inesperado y 'estadísticamente significativo' del riesgo de cáncer de pulmón (como se vio en el Ejemplo 7.17). Los autores consideraron posibles explicaciones alternativas para esta observación inesperada. La confusión se podía descartar debido al gran tamaño de la muestra y la asignación aleatoria de los sujetos a los diferentes grupos de estudio. Además, los grupos de tratamiento estaban bien equilibrados respecto a las características basales importantes. Como este resultado no estaba avalado por otros estudios biológicos o epidemiológicos, los autores no quisieron rechazar la hipótesis nula de ningún efecto. Al final de su artículo indicaron:

«En resumen, no encontramos una reducción global de la incidencia de cáncer de pulmón o de la mortalidad debida a esta enfermedad entre los fumadores varones que recibieron suplementos en la dieta de alfa-tocoferol, beta-caroteno, o ambos en este gran ensayo en Finlandia. Los resultados de este estudio sugieren que estas sustancias pueden ejercer efectos dañinos así como efectos beneficiosos. Para determinar el es-

pectro completo de los efectos de estos agentes sería necesario la observación durante más tiempo de los participantes en este estudio y la comparación con datos de otros estudios en personas con riesgo normal o alto de padecer cáncer. En estos momentos sería prematuro proponer recomendaciones de salud pública en relación a los suplementos dietéticos con estos micronutrientes» (*The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group*, 1994).

Más adelante se publicaron los resultados de otros dos ensayos. La fase de intervención activa del ensayo CARET (Ejemplo 7.14) se terminó prematuramente tras confirmar el aumento inesperado del riesgo de cáncer de pulmón entre aquellos que tomaban beta-caroteno al igual que el observado en el ensayo ATBC. Tampoco había una explicación obvia para este hallazgo inesperado. El estudio sobre la salud de los médicos (Ejemplo 7.8) tenía una duración de seguimiento mucho mayor (12 años) que cualquiera de los otros ensayos (media de 6 y 4 años para los ensayos ATBC y CARET, respectivamente). Sus resultados eran compatibles con la hipótesis nula según la cual no había ningún efecto, beneficioso o dañino, de los suplementos de beta-caroteno sobre el riesgo de cáncer de pulmón. Por tanto, la historia de beta-caroteno y cáncer de pulmón muestra que los resultados de un ensayo único no deben considerarse de forma aislada.

También deben tenerse en cuenta otros factores antes de que los resultados de un ensayo puedan traducirse en decisiones de salud pública. Éstos incluyen aspectos tales como la posibilidad de generalizar sus resultados a otras poblaciones, la aceptabilidad de la intervención, factibilidad, costes, recursos disponibles y prioridades competitivas en salud pública. Cabe destacar que el impacto global de la intervención en una población en particular depende, no sólo de la magnitud del efecto de la intervención sobre el riesgo de desarrollar una enfermedad en particular, sino también de la frecuencia (y gravedad) de la enfermedad en cada población. Este aspecto se comentará más ampliamente en el Capítulo 16.

Lecturas adicionales

* El libro Smith y Morris (1996) proporciona una visión completa del diseño, realización, y monitorización de los ensayos de campo con un énfasis especial en los aspectos prácticos. Aunque se destacan los países en desarrollo, la mayoría de los aspectos comentados en este libro también son importantes para países desarrollados.

* El libro de Pocock (1983) proporciona una referencia buena y accesible para los interesados en el diseño, análisis e interpretación de los ensayos clínicos.

* En dos artículos de Peto *et al.* (1976, 1977) puede encontrarse una breve revisión de los aspectos metodológicos del diseño, análisis e interpretación de los ensayos clínicos en temas de cáncer.

Cuadro 7.1. Puntos esenciales

- Los estudios de intervención se caracterizan por el hecho de que los investigadores son responsables de asignar los pacientes a los diferentes grupos del estudio.
- Las principales ventajas de este tipo de estudio son:
 1. La *asignación aleatoria* de los sujetos asegura que la asignación de los sujetos a los diferentes grupos del estudio no esté afectada por sesgos de selección.
 2. La *asignación aleatoria* asegura que los grupos sean equilibrados respecto a los factores conocidos y, más importante, los desconocidos que puedan afectar al resultado del estudio (siempre que el estudio sea lo suficientemente grande).
 3. Si la asignación es a *doble ciego* también se reducen los sesgos de medida.
 4. Para una intervención pueden estudiarse *múltiples resultados*.
 5. La incidencia de la enfermedad puede medirse en los diferentes grupos de estudio.
- Los principales inconvenientes de este tipo de estudio son:
 1. Los estudios de intervención, especialmente los estudios de campo, son empresas grandes. Son muy caras y requieren mucho tiempo.
 2. Pueden presentar problemas éticos importantes.
 3. Puede ser difícil asegurar el cumplimiento del régimen de intervención y evitar la contaminación entre los grupos en el transcurso del estudio, especialmente en estudios de larga duración.

Anexo 7.1

Tabla de números aleatorios

16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62
56 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	75 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43

Tabla A7.1.
Tabla de números aleatorios (tabla
XXXIII de Fischer y Yates (1963)).

Capítulo 8

Estudios de cohorte

Un estudio de cohorte es un estudio observacional en el que se selecciona una población, o cohorte, y se obtiene información para determinar qué sujetos tienen una característica particular (p.e. grupo sanguíneo A) que se sospecha está relacionada con el desarrollo de la enfermedad investigada o han estado expuestos a un posible agente etiológico (p.e. fumar cigarrillos). Toda la población en estudio se sigue en el tiempo y se compara la incidencia de la enfermedad en los individuos expuestos con la incidencia en los no expuestos.

Así, los estudios de cohorte se parecen a estudios de intervención en que las personas se seleccionan basándose en su estado de exposición y después se siguen en el tiempo, pero difieren en que la asignación a los grupos de estudio no está bajo el control directo de los investigadores.

***Ejemplo 8.1.** Se planteó un estudio de cohorte de 22.707 hombres chinos en Taiwan para investigar la asociación entre el antígeno de superficie de la hepatitis B (HBsAg) y el desarrollo de carcinoma hepatocelular. El estudio se realizó en empleados del gobierno chino incluidos a partir de las revisiones médicas rutinarias. Todos los participantes cumplieron un cuestionario de salud y proporcionaron una muestra de sangre a la inclusión del estudio. Los participantes se siguieron durante una media de 3,3 años (Beasley et al., 1981).*

En el Ejemplo 8.1, se reunió un grupo de 22.707 hombres empleados del gobierno chino (la 'cohorte') y se determinó la seropositividad del HBsAg (la 'exposición') al inicio del estudio. A continuación se siguieron en el tiempo para determinar (y comparar) la incidencia de carcinoma hepatocelular (el 'resultado') en sujetos HBsAg positivos y HBsAg negativos en el momento de la inclusión en el estudio.

8.1 Definición clara de los objetivos

Al igual que con cualquier otro diseño de estudio, es esencial que se formule una hipótesis clara antes del inicio del estudio. Esto debe incluir una definición clara de la(s) exposición(es) y resultado(s) de interés. Debido a que los estudios de cohorte en epidemiología del cáncer con frecuencia implican el seguimiento de un gran número de personas durante un periodo de tiempo prolongado, tienden a ser muy caros y requerir mucho tiempo. En consecuencia, estos estudios generalmente se efectúan después de que

la hipótesis se haya explorado en otros tipos de estudios más baratos y rápidos (p.e. estudios transversales o caso-control). Por ejemplo, el estudio de cohorte en Taiwan (Ejemplo 8.1) se inició sólo después de que se hubieran realizado una serie de estudios caso-control de carcinoma hepatocelular a principios y mediados de los años 70 (IARC, 1994b).

8.2 Elección de la población del estudio

8.2.1 Fuentes de la población del estudio

La elección de un grupo de individuos para un estudio de cohorte determinado depende de las hipótesis específicas que quieren investigarse y de las limitaciones prácticas para la realización del estudio. La cohorte elegida puede ser un grupo de la población general, como los residentes de una comunidad o una población más específica que pueda identificarse y seguirse directamente, tales como los miembros de organizaciones profesionales o sociales (p.e. miembros de seguros sanitarios, registros de médicos y enfermeras, etc.). De manera alternativa, la cohorte puede elegirse debido a una exposición elevada a un factor etiológico sospechado, tales como una fuente de radiación ionizante, un tipo particular de tratamiento (p.e. quimioterapia, radioterapia) o un riesgo ocupacional.

Una *cohorte de la población general* puede reclutarse de una zona geográfica bien definida (como en el Ejemplo 8.2), para establecer el estado de exposición basal respecto a un número de factores y después se examinan periódicamente para evaluar futuros resultados.

Ejemplo 8.2. Una cohorte de 10.287 individuos residentes en el distrito de Ernakulam de Kerala (India) se siguió durante un periodo de 10 años para evaluar el efecto del hábito de mascar tabaco sobre la mortalidad global. Los participantes se identificaron inicialmente mediante un censo en el que se seleccionó un número de pueblos del distrito. Se entrevistó a todos los residentes mayores de 15 años de los pueblos elegidos sobre sus hábitos de mascar tabaco mediante una entrevista puerta a puerta y se incluyeron en la cohorte. Las negativas a participar fueron muy pocas (Gupta et al., 1984).

Una de las grandes ventajas de este tipo de estudios de cohorte es que permite evaluar un gran número de exposiciones comunes en relación a un gran número de resultados. El estudio de Framingham es un ejemplo clásico de esto. Aproximadamente 5.000 residentes de la localidad de Framingham, en Massachusetts (EE.UU.), se siguieron desde 1948 (Dawber *et al.*, 1951). Había varias razones para elegir esta zona geográfica, principalmente consideraciones logísticas y prácticas para asegurar que sería factible la identificación y el seguimiento de los participantes durante muchos años. En el momento del inicio del estudio, Framingham era una comunidad relativamente estable, con áreas industriales y rurales, y tenía representada una gran variedad de industrias y ocupaciones. La localidad era lo suficientemente pequeña como para permitir que los residentes fueran a un único

centro de salud para los exámenes médicos y sólo había un hospital importante. Esta cohorte ha permitido la evaluación de los efectos de una gran variedad de exposiciones comunes (p.e. presión sanguínea, colesterol sérico, consumo de alcohol, ejercicio físico, fumar, etc.) sobre el riesgo de numerosas enfermedades, incluyendo las enfermedades cardiovasculares, cáncer, gota, patologías oculares, de la vesícula biliar, etc.

Como alternativa, por razones logísticas puede ser preferible tomar la cohorte de la población general a partir de un grupo socio-profesional de individuos bien definido. Por ejemplo, el estudio de Taiwan descrito en el Ejemplo 8.1 no se realizó entre empleados del gobierno porque se creía que tendrían una mayor exposición al virus de la hepatitis B que el resto de la población, sino porque este grupo de personas era más fácil de identificar y seguir que cualquier otra población potencial de estudio.

Ejemplo 8.3. *En 1951 se envió un cuestionario por correo a todos los hombres y mujeres cuyos nombres se encontraban en aquel tiempo en el Registro de Médicos Británicos [British Medical Register] y que se creía eran residentes del Reino Unido. Además del nombre, dirección y edad, se les preguntó algunas cuestiones simples sobre su hábito tabáquico. Un total de 33.439 hombres y 6.194 mujeres médicos proporcionaron respuestas lo suficientemente completas. Estos médicos se han seguido desde entonces. (Doll y Hill, 1954; Doll y Peto, 1976; Doll et al., 1980, 1994a,b).*

Asimismo, cuando Richard Doll y Bradford Hill decidieron iniciar un estudio de cohorte en Inglaterra y Gales para evaluar los efectos sobre la salud del tabaco, se decidió elegir a los médicos británicos como población de estudio (Ejemplo 8.3), principalmente por razones logísticas. Los médicos estaban registrados en la Asociación Británica de Médicos [British Medical Association] y por tanto, eran fáciles de identificar y seguir. Además, era más probable que cooperaran y que se diagnosticara correctamente la causa de muerte.

Si la *exposición es poco común*, un estudio de la población general tendrá poca capacidad para detectar un efecto (es decir, el estudio no tendrá el suficiente poder estadístico (véase Capítulo 15)) ya que muy pocas personas habrán estado expuestas al factor de interés. Este problema puede resolverse seleccionando intencionadamente como población de estudio a un *grupo de personas con exposiciones muy altas*. Por ejemplo, la exposición a tintes y colorantes es rara en la población general. No obstante, eligiendo un grupo de trabajadores con alta exposición, podrá estudiarse potencialmente todo el espectro de efectos, incluyendo aquellos que son poco frecuentes en la población general pero no en los grupos muy expuestos. El impacto general de la exposición sobre la salud pública puede ser bajo, pero estos estudios pueden proporcionar una valiosa información sobre los mecanismos biológicos comunes de la historia natural de enfermedad.

El seguimiento de los supervivientes de las explosiones de las bombas atómicas en Japón (Ejemplo 8.4) no sólo ha aclarado muchos de los efectos

Ejemplo 8.4. *El estudio sobre la duración de la vida [The Life Span Study] es un estudio de cohorte en curso que se inició para investigar los efectos sobre la salud a largo plazo de la exposición a altos niveles de radiación ionizante entre los supervivientes de las explosiones de las bombas atómicas en Japón. Cuando se inició el seguimiento incluía una muestra de 120.128 sujetos residentes de Hiroshima y Nagasaki en 1950 (Shimizu et al., 1990).*

a largo plazo de la exposición aguda a niveles altos de radiación ionizante sino que también ha contribuido a entender los efectos de la exposición crónica a niveles bajos de radiación.

8.2.2. Elección del grupo de comparación

Una vez determinada la fuente de los sujetos expuestos, la siguiente fase es elegir un grupo de comparación adecuado de individuos no expuestos. Esta selección es el aspecto más crítico en el diseño de un estudio de cohorte. El grupo no expuesto debe ser lo más parecido posible al grupo expuesto respecto a la distribución de todos los factores que puedan estar relacionados con el (los) resultado(s) de interés a excepción de la exposición en estudio. En otras palabras, si realmente no hubiera ninguna asociación entre la exposición y la enfermedad, las tasas de enfermedad en los dos grupos comparados serían esencialmente las mismas. En un estudio de cohorte pueden utilizarse dos tipos principales de grupos de comparación: *internos* y *externos*.

Las cohortes de la población general tienden a ser heterogéneas respecto a muchas exposiciones y, por tanto, sus miembros pueden clasificarse en diferentes categorías de exposición. En estas circunstancias, puede utilizarse un grupo interno de comparación. Esto es, la experiencia de aquellos miembros de la cohorte que no están expuestos o los que están expuestos a niveles bajos pueden utilizarse como grupo de comparación. Por ejemplo, en el estudio de cohorte de los médicos británicos (Ejemplo 8.3), fue posible clasificar a los individuos en categorías en función del hábito tabáquico

Ejemplo 8.5. *El Estudio sobre la Salud de las Enfermeras [Nurses' Health Study] es un estudio de cohorte en curso iniciado en 1976, cuando 121.700 enfermeras registradas en los EE.UU. de edades comprendidas entre 30 y 55 años cumplimentaron un cuestionario sobre su estado de salud y estilo de vida. Un total de 1.799 casos nuevos de cáncer de mama se diagnosticaron durante los 10 primeros años de seguimiento desde mediados de 1976 a mediados de 1986. A continuación se realizaron análisis estadísticos para determinar la relación entre el uso de anticonceptivos orales y el riesgo de cáncer de mama. Las mujeres que contestaron que nunca habían tomado anticonceptivos orales en el cuestionario de 1976 y en los sucesivos se consideraron como el grupo de 'no expuestos' (Romieu et al., 1989).*

y después comparar la mortalidad por cáncer de pulmón (y otras causas) entre los fumadores y los no fumadores.

En el Ejemplo 8.5, una parte de la cohorte de enfermeras registradas en los EE.UU. se consideraron como el grupo de 'no expuestas' para estudiar la relación entre el uso de anticonceptivos orales y el riesgo de desarrollar cáncer de mama.

En las cohortes de población general puede examinarse el efecto de más de una exposición. Por tanto, la elección del grupo de personas en la cohorte que se considerará como 'no expuesto' depende de cada exposición particular en estudio. Por ejemplo, el Estudio sobre la Salud de las Enfermeras también ha permitido evaluar, entre otros, la relación entre la ingesta total de grasa en la dieta y el riesgo de cáncer de mama. En este caso, se determinó la distribución de la ingesta de grasa en toda la cohorte y ésta se dividió en cinco grupos de igual tamaño ('quintiles'). Las mujeres en el quintil inferior se tomaron como el grupo de 'no expuestos' (Willett *et al.*, 1987).

En cohortes ocupacionales, un grupo interno de comparación estaría formado por los trabajadores empleados en otras ocupaciones dentro de la misma industria que no incluyeran la exposición al factor en estudio.

Ejemplo 8.6. *El 31 de diciembre de 1975 se identificó retrospectivamente, a partir de registros ocupacionales, una cohorte de 14.282 trabajadores de la planta de Sellafield de Combustibles Nucleares Británicos [British Nuclear Fuels]. Todos ellos trabajaban desde la apertura de la planta en 1947. Los empleados que trabajaron en zonas en las que era posible que estuvieran expuestos a radiación externa llevaron dosímetros de placa y la industria mantuvo los registros individuales de dosis. Estos trabajadores se consideraron en el estudio como 'trabajadores con radiación' mientras que los que nunca llevaron dosímetros se consideraron 'trabajadores sin radiación'. Inicialmente se planeó seguir a los trabajadores desde el momento en que empezaron a trabajar hasta finales de 1975, pero el periodo de seguimiento se amplió más adelante hasta finales de 1988. La mortalidad experimentada por los trabajadores 'con radiación' se comparó entonces con la experimentada por los trabajadores 'sin radiación' (Smith y Douglas, 1986; Douglas *et al.*, 1994).*

En el Ejemplo 8.6, fue posible identificar a un grupo de trabajadores que podía considerarse como no expuesto a radiaciones externas basándose en los registros individuales de dosis.

Cuando la cohorte es esencialmente homogénea en relación a la exposición al factor sospechado, es necesaria una cohorte similar pero no expuesta o algún otro estándar de comparación para evaluar la experiencia del grupo expuesto. Por ejemplo, en algunas cohortes ocupacionales, con frecuencia no es posible identificar un subgrupo que pueda considerarse 'no expuesto'. En este caso, se tendrá que utilizar un *grupo de comparación externo*. Un grupo de comparación potencial es una cohorte de trabajadores similares incluidos en otra ocupación que no implica la exposición al

factor de interés. Por ejemplo, muchas exposiciones sólo se producen entre ciertas ocupaciones y por tanto, puede suponerse que el nivel de exposición de otros trabajadores en el resto de ocupaciones es virtualmente cero. Por tanto, podemos elegir como grupo control a personas con empleo y residentes en la misma zona geográfica que no están expuestas al mismo factor de riesgo. Es importante asegurar que el riesgo de la enfermedad en estos trabajadores no está afectado por sus propias exposiciones ocupacionales.

Como alternativa, la *población general* de la zona geográfica en la que residen los individuos expuestos puede tomarse como grupo de comparación externo. En este caso, la experiencia de enfermedad observada en la cohorte se compara con la experiencia de enfermedad de la población general en el tiempo en que se sigue la cohorte. La comparación con las tasas en la población general evita la necesidad de tener que seguir a un gran número de individuos no expuestos, pero presenta varios inconvenientes. En primer lugar, sólo puede realizarse para enfermedades para las cuales exista esta información para la población general. En segundo lugar, presupone que sólo una pequeña proporción de la población general está expuesta al factor de riesgo de interés, de lo contrario, el efecto real de la exposición estará infraestimado ya que la exposición y sus efectos están presentes también en el grupo de comparación. En tercer lugar, incluso si se elige la población general lo más parecida posible a la cohorte expuesta, en lo que respecta a las características demográficas y geográficas básicas, pueden ser muy diferentes respecto a otros factores de riesgo para la enfermedad tales como la dieta, tabaco, etc. Debido a que esta información no está disponible para los individuos de una población general, las diferencias observadas en nuestro estudio pueden ser debidas a los efectos de factores de confusión que no se han podido controlar.

La ventaja de utilizar otro grupo de personas como grupo externo de comparación en lugar de la población general, es que puede elegirse el grupo de modo que sea más comparable a la cohorte expuesta de lo que lo sería la población general. Además, puede obtenerse información sobre factores potenciales de confusión de todos los individuos expuestos y no expuestos en el estudio y controlar las diferencias en el análisis.

En muchos estudios de cohortes puede ser útil tener *múltiples grupos de comparación*, especialmente cuando no puede asegurarse que ninguno de los grupos sea lo suficientemente similar al grupo expuesto en términos de distribución de las variables de confusión. En estas circunstancias, los resultados del estudio pueden ser más convincentes si se observa una asociación similar para grupos de comparación diferentes. Por ejemplo, en algunas cohortes ocupacionales puede utilizarse un grupo de comparación interno (personas empleadas en la misma fábrica pero con un trabajo diferente) y la experiencia de la población general (tasas nacionales y locales).

En el Ejemplo 8.7, la mortalidad global de la cohorte de trabajadores del caucho se comparó con la mortalidad de otra cohorte industrial y con las

Ejemplo 8.7. Se determinó una cohorte de trabajadores de una planta de fabricación de neumáticos en Akron, Ohio (EE.UU.) para estudiar la mortalidad global y específica. El 1 de enero de 1964 se identificaron retrospectivamente, a partir de archivos de pensionistas, nóminas, certificados de defunción y otros, un total de 6.678 hombres de 40 a 84 años trabajadores de la industria del caucho. Estos trabajadores se siguieron de 1964 a 1972. La mortalidad específica por edad de esta cohorte se comparó entonces con la de otros tres grupos de comparación, una cohorte industrial de trabajadores del acero, la población del estado en la que se encuentra la planta (Ohio) y la población nacional de los EE.UU. (Tabla 8.1) (McMichael et al., 1974).

Grupo de edad (años) ^b	Tasas de mortalidad específica (por 100.000 p-as)			
	Cohorte de trabajadores del caucho (1964-72)	Cohorte de trabajadores del acero (1953-61)	Estado de Ohio (1972)	EE.UU. (1968)
45-54	852	907	940	980
55-64	2.317	2.166	2.365	2.370

^a Datos de McMichael et al. (1974).
^b Sólo se dispone de datos para estos dos grupos de edad para las cuatro poblaciones.

tasas locales (del estado) y nacionales. Observe que los trabajadores del caucho y del acero experimentaron unas tasas de mortalidad específicas por edad menores que las poblaciones del estado o nacional. Esto es debido a que las personas que trabajan tienden a ser más sanas que la población general, ya que ésta incluye a los que están demasiado enfermos o incapacitados para trabajar (aunque para los trabajadores del acero la diferencia puede ser debida en parte a cambios en la mortalidad con el tiempo). Este sesgo de selección se denomina el 'efecto del trabajador sano'.

El efecto del trabajador sano puede enmascarar aumentos reales del riesgo de una enfermedad en relación a una exposición en particular. Se sabe que la importancia de este sesgo varía con el tipo de enfermedad, siendo menor para el cáncer que para otras enfermedades y tiende a desaparecer con el tiempo a partir del primer contrato en una ocupación (véase Apartado 13.1.1). Si las tasas permanecen inferiores a las de la población general durante todo el seguimiento, es más probable que sea debido a diferencias socio-demográficas y del estilo de vida entre los trabajadores y la población general que a la selección de individuos sanos en el momento de la inclusión.

Tabla 8.1.

Tasas de mortalidad específicas por edad en la cohorte de trabajadores del caucho y en los otros tres grupos de comparación: trabajadores del acero, población del estado de Ohio y población nacional de los EE.UU.^a

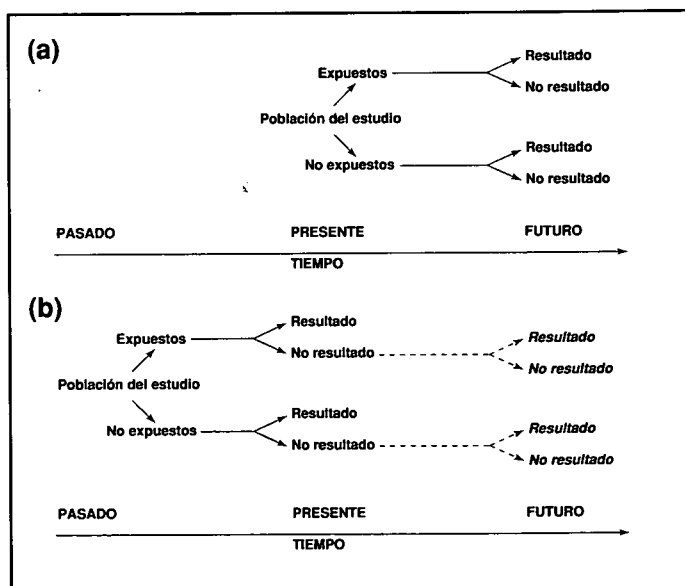


Figura 8.1.

Perfil de un estudio de cohorte prospectivo (a) y de uno histórico (b).

Este efecto de la selección del trabajador sano no está limitado a cohortes ocupacionales. Se ha observado un fenómeno similar en muchos otros tipos de estudios de cohorte. En el estudio de los médicos británicos descrito en el Ejemplo 8.3, los que contestaron al cuestionario en los primeros años de seguimiento presentaron una mortalidad menor que los que no contestaron (Doll y Hill, 1954). Las personas menos preocupadas por la salud, o aquellas que ya padecían problemas de salud, podían haber estado menos motivadas a participar.

8.3 Medida de la exposición

La medida de la(s) exposición(es) de interés es un aspecto crucial en el diseño de un estudio de cohorte. Debe obtenerse información sobre la edad al inicio de la exposición, fechas en que se inició y finalizó la exposición, dosis y patrón de exposición (intermitente frente a constante) y cambios en el tiempo (véase Apartado 2.3).

La información sobre la(s) exposición(es) de interés puede obtenerse a partir de diferentes fuentes, incluyendo registros independientes del estudio (p.e. registros médicos, ocupacionales o sindicales); información suministrada por los propios sujetos del estudio, mediante entrevistas o cuestionarios; datos obtenidos por examen médico u otro control de los participantes; muestras biológicas o medidas directas del ambiente en que han vivido o trabajado los miembros de la cohorte. Las ventajas y limitaciones de cada una de estas fuentes se comentaron en el Capítulo 2.

Existen dos tipos principales de estudios de cohorte definidos según el momento en el tiempo en que se obtuvo la información sobre la exposición: presente o pasado. En *estudios de cohorte prospectivos*, los datos sobre la exposición se obtienen ahora, una vez se ha definido la población en estudio. En este caso, pueden utilizarse los métodos más modernos de medida de la exposición de modo que el sesgo en la clasificación de la exposición puede reducirse al máximo. El principal inconveniente de este tipo de estudios de cohorte es que el tiempo desde la exposición hasta el inicio de la enfermedad (es decir, el periodo de inducción) puede ser demasiado largo (muchas décadas para la mayoría de los cánceres). Los Ejemplos 8.1 a 8.5 son ejemplos de estudios de cohorte prospectivos que implicaban el seguimiento de un gran número de personas durante periodos de tiempo muy prolongados.

Una alternativa, especialmente útil para enfermedades con periodos de inducción muy largos, es confiar en las medidas de la exposición efectuadas muchos años antes del inicio del estudio y que pueden obtenerse a partir de registros médicos, ocupacionales u otro tipo de registros. Utilizando datos de registros preexistentes, el tiempo que tendríamos que esperar para que la exposición tuviera un efecto sobre el riesgo de padecer la enfermedad se reduciría considerablemente. Este tipo de estudio de cohorte se denominan *estudios de cohorte históricos*.

El estudio descrito en el Ejemplo 8.8 es un ejemplo clásico de este tipo de diseño. Los Ejemplos 8.6 y 8.7 también son ilustraciones de estudios de co-

Ejemplo 8.8. *A principios de los años 50, Case y sus colaboradores iniciaron un estudio de cohorte para evaluar si los trabajadores en la fabricación de determinadas sustancias intermediarias de colorantes tenían un exceso de riesgo de cáncer de vejiga. Realizaron una lista de todos los hombres que habían estado empleados como mínimo 6 meses desde 1920 en la industria química en el Reino Unido. Registraron la edad y las fechas entre las que se produjo la exposición a los colorantes. Se realizó una búsqueda retrospectiva de todos los casos de cáncer de vejiga aparecidos en los empleados de la industria química, desde 1921 hasta febrero de 1952. El número de casos de cáncer de vejiga observados entre estos trabajadores se comparó entonces con el número esperado si estos trabajadores tuvieran la misma mortalidad que la población general del Reino Unido (Case et al., 1954; Case y Pearson, 1954).*

horte históricos ya que los dos se basan en registros de empleo preexistentes para identificar los miembros de la cohorte y clasificarlos según su exposición. Los estudios de cohorte históricos son especialmente útiles en la epidemiología ocupacional porque si se sospecha que una exposición particular puede ser de alto riesgo para una enfermedad, no es razonable esperar durante décadas a que esto se aclare en un estudio de cohorte prospectivo. No obstante, si en el momento que se identifica la cohorte histórica todavía están vivos una gran proporción de sus miembros, el periodo de seguimiento puede ampliarse en el futuro (como en el Ejemplo 8.6) para asegurar la evaluación de todos los posibles efectos a largo plazo.

Una de las principales limitaciones de los estudios de cohorte históricos es que los datos disponibles de exposición en los registros del pasado suelen ser menos exactos y detallados que si se recopilan de manera prospectiva. Así por ejemplo, en la mayoría de las cohortes ocupacionales, raramente se dispone de las determinaciones de exposiciones pasadas efectuadas por medición directa en el ambiente de trabajo, y por tanto las variables tales como la asignación del trabajo o el ser miembro de un sindicato o asociación profesional, se utilizan generalmente para clasificar la exposición individual. Estas variables de aproximación son, en el mejor de los casos, sólo marcadores crudos de los niveles reales de exposición y los detalles disponibles pueden no ser suficientes para tratar correctamente las cuestiones específicas investigadas. No obstante, es poco probable que la exactitud o la exhaustividad de estos registros sea diferente para los que desarrollaron la enfermedad de interés y los que no la desarrollaron ya que los datos se registraron antes de producirse el evento, y en la mayoría de los casos, por razones completamente independientes de la investigación. Siempre que el error de clasificación de exposición sea independiente del resultado, es decir, no sea diferencial, tenderá a diluir toda asociación real entre la exposición y el resultado (véanse Apartados 2.7 y 13.1.2).

Un aspecto especialmente interesante de las cohortes históricas es cuando en el pasado se conservaron muestras biológicas que pueden analizarse actualmente con técnicas de laboratorio más sensibles para medir la expo-

sición. Por ejemplo, acceder a los bancos de suero para determinar la exposición a agentes infecciosos (como en el Ejemplo 8.9) o a sustancias químicas. Este método reduce inexactitudes en la determinación de la exposición en el pasado, pero el número de estos bancos de muestras biológicas es limitado y muchas veces no se conoce la estabilidad del marcador biológico durante periodos largos de conservación.

***Ejemplo 8.9.** Muchos estudios habían notificado títulos elevados de anticuerpos IgG contra el antígeno de la cápside y una alta prevalencia de anticuerpos contra el antígeno temprano del virus Epstein-Barr en pacientes con la enfermedad de Hodgkin. No obstante, en estos estudios, las muestras de sangre se habían tomado tras el diagnóstico y tratamiento de la enfermedad de Hodgkin. Para evaluar si la activación del virus Epstein-Barr precedía al desarrollo de la enfermedad de Hodgkin se realizó un estudio en colaboración con cinco bancos de suero localizados en los EE.UU., Noruega y Suecia. Se disponía de muestras de suero de más de 240.000 personas. Los pacientes a los que más adelante se les diagnosticó cáncer, se identificaron mediante la búsqueda nominal en los archivos de los hospitales y en los registros de cáncer. Se identificaron 43 casos de enfermedad de Hodgkin y se analizó la presencia del virus Epstein-Barr en los sueros almacenados (Mueller et al., 1989).*

En muchos estudios de cohorte se efectúa una única clasificación de cada individuo según su exposición en el momento de su inclusión en el estudio. Esto es adecuado para las exposiciones que no cambian con el tiempo. No obstante, durante el seguimiento a largo plazo, con frecuencia se producirán cambios en los niveles de la exposición para los factores de interés. Los individuos pueden cambiar de trabajo, decidir dejar de fumar (como en el Ejemplo 8.10) o adoptar una dieta pobre en grasas. En muchos casos, es posible repetir las determinaciones de la exposición inicial durante el periodo de seguimiento (o puede disponerse de la información de estos cambios por registros históricos). Así el riesgo de desarrollar la enfermedad puede estudiarse en relación al estado de exposición inicial y a los cambios subsiguientes.

Puede haber otras razones para volver a evaluar el estado de exposición de los sujetos del estudio, especialmente en estudios prospectivos a largo plazo. En el curso del estudio pueden aparecer métodos más actualizados para la determinación de las exposiciones o información científica nueva sobre la enfermedad que pueden crear la necesidad (o el interés) de medir variables adicionales que no se determinaron inicialmente.

En el Ejemplo 8.10, durante el periodo de seguimiento de 40 años, se observaron unos cambios importantes en el hábito de fumar de los médicos varones británicos (Tabla 8.2). Un 62% de los médicos manifestaron ser fumadores en 1951. La cifra correspondiente en 1990 fue sólo del 18%. Estos cambios en el estado de la exposición pueden tenerse en cuenta en el análisis de estudios de cohorte (Apartado 8.7).

Ejemplo 8.10. En el estudio de los médicos británicos descrito en el Ejemplo 8.3, se envió un primer cuestionario a todos los médicos británicos registrados en 1951. Se realizaron más indagaciones sobre los cambios en el hábito de fumar en 1957, 1966, 1972, 1978 y 1990-91. Los dos últimos cuestionarios también incluyeron preguntas adicionales sobre el consumo de alcohol y otras características personales. Para evaluar la magnitud de los cambios en el hábito de fumar durante el periodo de seguimiento de 40 años, se compararon las características del tabaquismo de los hombres que contestaron al cuestionario de 1990-91 con lo que habían indicado en el cuestionario de 1951 (Tabla 8.2) (Doll et al., 1994a).

Hábito tabáquico 1951	Hábito tabáquico, 1990-91					Nº (%) en 1951
	No fumador	Ex fumador	Fumador actual			
			Sólo cigarrillos	Cigarrillos y otros	Pipa o cigarros	
No fumador	2.361 ^a	198	17	4	86	2.666 (25)
Ex fumador	0	1.374	10	3	66	1.453 (13)
Fumador actual						
Sólo cigarrillos	0	3.355	535	47	446	4.383 (41)
Cigarrillos y otros	0	897	74	31	308	1.310 (12)
Pipa o cigarros	0	695	16	2	287	1.000 (9)
Nº (%) en 1990-91	2.361 (22)	6.519 (60)	652 (6)	87 (1)	1.193 (11)	10.812 (100)

^a Datos de Doll et al. (1994a).

Tabla 8.2.

Hábitos de fumar de los hombres participantes en el estudio de médicos británicos que contestaron a los cuestionarios de 1951 y 1990-91^a.

8.4 Medida de las variables de confusión potenciales

Los estudios de cohorte son estudios observacionales y por tanto los individuos no se distribuyen aleatoriamente a las diferentes categorías de exposición. Esto puede dar lugar a diferencias entre los grupos en relación a exposiciones diferentes a la estudiada. Esto sólo es importante si estas otras exposiciones también son factores de riesgo de la enfermedad en particular (u otro resultado) en estudio, es decir, si estas exposiciones son variables de confusión. De este modo, si se está estudiando una exposición ocupacional y cáncer de pulmón, es necesario estar seguro que los grupos 'expuesto' y 'no expuesto' presentan una historia similar del hábito tabáquico. Si no es así, tendrá que hacerse un *ajuste estadístico* de las diferencias en los hábitos de fumar (véanse Capítulos 13-14). Con objeto de estar seguros de que no hay diferencias, o para efectuar este ajuste, son necesarios los datos sobre el consumo de tabaco de cada individuo. Estos datos deben ser lo más exactos posible y de calidad similar a los recogidos sobre la exposición de interés objeto de estudio.

En estudios de cohorte históricos, con frecuencia falta la información sobre factores de confusión. Ésta es una de las principales limitaciones. Por ejemplo, en muchas de las cohortes ocupacionales históricas establecidas para investigar la relación entre la exposición al amianto y cánceres respi-

ratorios, no se disponía de información sobre el hábito de fumar. Por el contrario, la recogida de datos sobre los factores de confusión puede incluirse en el diseño de la mayoría de los estudios de cohorte prospectivos, a excepción de cuando se toman tasas de la población general locales o nacionales como grupo de comparación no expuesto.

8.5 Medida del (los) resultado(s)

Una ventaja considerable de los estudios de cohorte es que puede examinarse el efecto de una exposición en particular sobre *múltiples enfermedades o resultados*.

Muchos estudios de cohorte utilizan los sistemas de vigilancia rutinarios disponibles para evaluar los resultados de interés. Estos sistemas pueden ser registros del cáncer, certificados de defunción o algún otro método de vigilancia especializado. La principal ventaja de utilizar estos sistemas para seguir a los sujetos del estudio y evaluar sus resultados es que puede realizarse a un coste mucho menor y sin que el investigador tenga que ponerse en contacto personalmente con los sujetos. El inconveniente es que estamos limitados a evaluar el tipo de resultados que se registran sistemáticamente y de acuerdo al modo en que estén codificados. Esto es especialmente importante en cohortes que duran varias décadas, debido a que pueden introducirse cambios sustanciales durante el periodo de estudio en el modo de identificación y codificación de las enfermedades por estos sistemas de vigilancia (véase Anexo A.2.2).

Cuando no se dispone de ningún sistema de vigilancia epidemiológica de enfermedades o cuando éstos no registran sistemáticamente la enfermedad o el resultado de interés, debe establecerse alguna forma de identificación y vigilancia de la enfermedad dentro de la cohorte. Por ejemplo, la evaluación de los resultados de interés puede realizarse mediante cuestionarios auto-administrados enviados regularmente a todos los sujetos del estudio, entrevistas personales, o mediante exámenes físicos regulares de todos los miembros de la cohorte.

Independientemente del método elegido para evaluar los resultados de interés, es vital asegurar que se utiliza de manera idéntica para los expuestos y los no expuestos. Si es posible, los entrevistadores y todas las personas implicadas en la evaluación de los resultados deberían desconocer la clasificación de exposición de los sujetos del estudio. De lo contrario, existe el potencial de introducir un sesgo de determinación (Apartado 13.1.2).

Los estudios de cohorte se centran en la aparición de la enfermedad. Para que se desarrolle una enfermedad, ésta debe estar, por supuesto, ausente en el momento de la inclusión de las personas en el estudio. Puede ser necesario un examen inicial de la población del estudio para identificar y excluir los casos existentes de enfermedad. Pero incluso cuando se realiza un examen basal, todavía puede ser imposible estar absolutamente seguro de que todos los individuos estaban exentos de la enfermedad al inicio del estudio, especialmente en enfermedades con un periodo de latencia largo (p.e. con un intervalo prolongado desde el inicio de la enfer-

medad hasta la aparición de los síntomas y signos clínicos). Por tanto, es habitual excluir los casos de enfermedad que se diagnostican durante un periodo de tiempo inmediatamente después de la inclusión en el estudio. Para el cáncer, se consideran habitualmente los 2-3 primeros años de seguimiento.

8.6 Seguimiento

Los *criterios de inclusión* en la cohorte deben definirse antes del inicio del estudio de manera clara e inequívoca. Los individuos sólo deben incluirse en la cohorte y contribuir al cómputo de persona-tiempo a riesgo, cuando se cumplen todos los criterios de inclusión. En la mayoría de los estudios de cohorte, los participantes se incluirán en la cohorte en diferentes momentos y por tanto, debe registrarse la fecha exacta de entrada de cada uno de ellos.

Al inicio del estudio, deben establecerse métodos para asegurar el seguimiento adecuado de los sujetos del estudio. En general, éstos incluyen contactos periódicos con los individuos tales como visitas domiciliarias, llamadas telefónicas, o cuestionarios por correo. Los estudios de cohorte de enfermedades con un periodo de latencia largo requieren el seguimiento de un gran número de sujetos durante muchos años. Esto es obviamente un trabajo muy grande y costoso. Para reducir estas dificultades, muchas cohortes se definen en términos de pertenencia a un grupo en particular (cuerpo profesional, plan de pensión ocupacional, sindicato, plan de seguro sanitario, alumnos de un colegio), en los que la población en estudio puede seguirse más fácilmente. En países en los que existen sistemas de vigilancia rutinarios, éstos pueden utilizarse para trazar y seguir los sujetos del estudio a un coste mucho menor que si los investigadores tuvieran que ponerse en contacto con ellos personalmente.

Los *criterios de exclusión* de la cohorte también deben definirse claramente. Debe especificarse una fecha como el final del periodo de seguimiento (como mínimo para el análisis actual). Por ejemplo, si la muerte es el resultado de interés, debe determinarse el estado vital en esta fecha para todos los miembros de la cohorte. Los sujetos cuyo estado vital se conoce en esta fecha, deben contribuir unidades de persona-tiempo a riesgo hasta esa fecha (o hasta la fecha de su muerte si se produce antes). Aquellos cuyo estado vital no se conoce en esta fecha deben considerarse 'perdidos en el seguimiento' y la última fecha en la cual se conoce su estado vital debe considerarse como el final de su contribución en unidades de persona-tiempo a riesgo.

Es esencial que se siga a la mayor proporción posible de personas de la cohorte. Algunas personas migrarán, morirán o cambiarán de empleo, pero deben realizarse todos los esfuerzos para determinar su estado vital y de salud. Todos estos factores pueden estar influenciados por la exposición y por tanto un seguimiento incompleto puede introducir un sesgo de selección (Apartado 13.1.1).

8.7 Análisis

El primer paso en el análisis de un estudio de cohorte es determinar la incidencia de la enfermedad (o de cualquier otro resultado de interés) en los expuestos y en los no expuestos, y compararlas.

Tabla 8.3.
Análisis de un estudio de cohorte: (a)
por riesgos (b) por tasas.

		Exposición	
		Sí	No
Resultado	Sí	<i>a</i>	<i>b</i>
	No	<i>c</i>	<i>d</i>
Riesgo en el grupo de expuestos (ρ_1) = $a/(a+c)$			
Riesgo en el grupo de no expuestos (ρ_0) = $b/(b+d)$			
Razón de riesgos = ρ_1/ρ_0			
Diferencia de riesgo ^a = $\rho_1 - \rho_0$			
		Exposición	
		Sí	No
Casos		<i>a</i>	<i>b</i>
Persona-tiempo a riesgo		<i>y</i> ₁	<i>y</i> ₀
Tasa en el grupo de expuestos (r_1) = a/y_1			
Tasa en el grupo de no expuestos (r_0) = b/y_0			
Razón de tasas = r_1/r_0			
Diferencia de tasas ^a = $r_1 - r_0$			
^a Si la exposición es protectora, las diferencias de riesgo y tasas deben calcularse como $\rho_0 - \rho_1$ o $r_0 - r_1$, respectivamente (véase Apartado 5.2.2)			

Si todos, o prácticamente todos los miembros de la cohorte se siguieron durante el mismo periodo de tiempo, podemos calcular el riesgo como la medida de aparición de la enfermedad en cada grupo (Tabla 8.3a). Por ejemplo, si el periodo es de 5 años uniformemente, puede calcularse el riesgo a 5 años por separado para los grupos de expuestos y de no expuestos. La razón de riesgos y la diferencia de riesgo pueden calcularse entonces como medidas de efecto relativo y absoluto, respectivamente.

Si los sujetos del estudio tienen periodos de seguimiento desiguales, deberán tenerse en cuenta en el análisis. Las duraciones del seguimiento pueden diferir considerablemente debido a que los sujetos se incluyeron en la población en estudio durante un periodo de tiempo relativamente largo, o porque se perdió el seguimiento de algunos antes de terminar el estudio (si, por ejemplo, cambian de residencia). Una manera de tratar periodos de seguimiento variables es calcular tasas, que utilizan *personas-años a riesgo* (o *personas-meses* o *personas-días*, etc., según sea más apropiado u oportuno) en el denominador (Tabla 8.3b). Con este método, cada sujeto contribuye a la población a riesgo sólo tantos años de observación como haya sido observado realmente; si el sujeto se retira después de 1 año, contribuye 1 persona-año; si se retira después de 10, habrá contribuido 10 personas-años (véase Apartado 4.2.2).

Las personas pueden contribuir personas-años de observación en más de un subgrupo. Suponga, por ejemplo, que en un estudio de 5 años, la incidencia de la enfermedad se determina para cada subgrupo de edad de 10 años. Una persona que se incluye en la cohorte al cumplir 48 años contribuirá 2 personas-años de observación al subgrupo de 40-49 años y 3 personas-años de observación al subgrupo de 50-59 años (véase Apartado 4.3.2). Esto también puede producirse con las determinaciones de la exposición si cambian en el tiempo. Por ejemplo, una persona puede ser un fumador durante unos años y después dejar de fumar.

Ejemplo 8.11. *El Estudio sobre la Salud de las Enfermeras ilustrado en el Ejemplo 8.5 consiste en una cohorte establecida a mediados de 1976 de 121.700 enfermeras registradas de los Estados Unidos de 30-55 años. Desde mediados de 1976 a mediados de 1986 se identificaron un total de 1.799 casos de cáncer de mama. Se realizaron análisis para investigar la relación entre el uso de anticonceptivos orales y el riesgo de cáncer de mama. Basándose en el cuestionario realizado a mediados de 1976, se planteó la siguiente pregunta: «si usted está usando o ha usado anticonceptivos orales, por favor indique los intervalos de uso de anticonceptivos orales empezando desde el primero y continuando hasta la actualidad. Si procede, por favor indique las razones para dejarlo». Se planteó la misma pregunta en los cuestionarios de seguimiento sucesivos a intervalos de dos años.*

En respuesta al cuestionario de 1976, 7.133 mujeres indicaron que estaban utilizando anticonceptivos orales. Las respuestas a los cuestionarios de 1978, 1980 y 1982 mostraron que 2.399, 1.168 y 302 mujeres, respectivamente, seguían tomando anticonceptivos orales. En 1984, ninguna de las mujeres los tomaban.

La información proporcionada en el cuestionario de 1976 se utilizó para clasificar a las enfermeras según las categorías de uso de anticonceptivos orales ('no usuarias', 'ex-usuarias' y 'usuarias actuales') y cada enfermera empezó a contribuir persona-tiempo a riesgo a estas categorías. Asimismo, para cada intervalo de 2 años subsiguiente, las mujeres contribuyeron el correspondiente persona-tiempo adicional a cada nuevo informe de uso de anticonceptivos orales. El seguimiento de las mujeres que desarrollaron cáncer de mama se truncó en el momento en que se les diagnosticó el cáncer (Romieu et al., 1989).

En el Ejemplo 8.11, las mujeres que cambiaron su situación en relación al uso de anticonceptivos orales durante el periodo de seguimiento habrían contribuido persona-tiempo a riesgo a diferentes categorías de exposición. Por ejemplo, una mujer que empezó utilizando anticonceptivos orales a principios de 1978 y paró al final de 1984 habría contribuido aproximadamente 1,5 personas-años a la categoría de no usuaria (desde el principio del estudio a mediados de 1976 al final de 1977), 7 personas-años a la catego-

ría de usuaria actual (desde el principio de 1978 al final de 1984), y 1,5 personas-años a la categoría de ex-usuaria (desde el principio de 1985 al final del seguimiento a mediados de 1986). Si esta mujer hubiera desarrollado cáncer de mama al final de 1982, su contribución persona-tiempo habría sido 1,5 personas-años a la categoría de no usuaria pero sólo 5 personas-años a la categoría de usuaria (su contribución persona-tiempo habría parado en el momento que desarrolló cáncer de mama).

Los resultados de interés también deben asignarse a alguna de las diferentes categorías de exposición. En nuestro ejemplo anterior, el caso de cáncer de mama debería asignarse a la categoría de usuaria de anticonceptivos orales ya que se produjo durante el tiempo en que la mujer estaba contribuyendo personas-años a esta categoría. Una vez asignados los personas-años a riesgo y los casos a las categorías de exposición correspondientes, pueden estimarse las tasas de incidencia de cáncer de mama para cada categoría de uso de anticonceptivos orales dividiendo el número de casos de cáncer de mama en cada categoría por el número total correspondiente de personas-años (Ejemplo 8.12).

Ejemplo 8.12. En el Ejemplo 8.11, se examinó la incidencia de cáncer de mama en enfermeras de 45-49 años en el momento de su inclusión en la cohorte, en relación al uso de anticonceptivos orales.

Tabla 8.4.

Distribución de casos de cáncer de mama y personas-años a riesgo en las enfermeras de los Estados Unidos de 45-49 años en el momento de su inclusión en la cohorte, según el uso de anticonceptivos orales^a.

	Uso de anticonceptivos orales		Total
	Usuarías alguna vez (actuales o en el pasado)	Nunca usuarias	
Casos	204	240	444
Personas-años a riesgo	94.029	128.528	222.557
Tasa por 100.000 p-as	217	187	199

^a Datos de Romieu *et al.* (1989)

Razón de tasas = 217 por 100.000 p-as/187 por 100.000 p-as = 1,16

Intervalo de confianza al 95% para la razón de tasas = 0,96 a 1,40

Diferencia de tasas = 217 por 100.000 p-as - 187 por 100.000 p-as = 30 por 100.000 p-as

Intervalo de confianza al 95% para la diferencia de tasas = - 8 a 68 por 100.000 p-as

$\chi^2 = 2,48$; 1 g.l.; $P=0,12$.

(Las pruebas estadísticas e intervalos de confianza se calcularon utilizando las fórmulas presentadas en el Anexo 6.1).

Los resultados del Ejemplo 8.12 muestran que no había diferencias estadísticamente significativas en la incidencia de cáncer de mama entre usuarias (las ex-usuarias y las usuarias actuales se agruparon en este análisis) y no usuarias de anticonceptivos orales.

La mayoría de las veces, las exposiciones en que estamos interesados pueden seguirse clasificando en diferentes niveles de intensidad. Los fu-

madores pueden clasificarse por el número de cigarrillos fumados al día, las usuarias de anticonceptivos orales por la duración total del uso, y las exposiciones ocupacionales por la intensidad de la exposición (con frecuencia estimada indirectamente a partir de datos sobre el tipo de trabajo o lugar de trabajo en la fábrica) o la duración del empleo. Si se utilizan diferentes niveles de exposición en la cohorte pueden examinarse las *tendencias* de la incidencia de la enfermedad según el nivel de exposición. Las conclusiones de un estudio se potencian si existe una tendencia a aumentar el riesgo (o disminuir, si la exposición es protectora) con niveles crecientes de exposición (es decir, si existe una relación *exposición-respuesta*).

Ejemplo 8.13. En el Ejemplo 8.11, también se examinó el riesgo de desarrollar cáncer de mama según el tiempo y la duración del uso de anticonceptivos orales (Tabla 8.5).

Tiempo Duración del uso	Casos	Personas- años	Tasa (por 100.000 p-as)	Razón de tasas (Intervalos de confianza al 95%)
No usuarias ^b	240	128.528	187	1,00
Usuarias actuales				
≤ 48 meses	4	328	1.220	6,52 (2,43–17,53)
> 48 meses	4	2.263	177	0,95 (0,35–2,55)
				test de χ^2 para tendencias = 0,46; $P = 0,50$
Usuarias en el pasado ^c				
≤ 48 meses	106	54.080	196	1,05 (0,84–1,32)
> 48 meses	86	36.039	239	1,28 (1,00–1,64)
				test de χ^2 para tendencias = 3,33; $P = 0,07$

^a Datos originales de Romieu *et al.* (1989).
^b Categoría basal.
^c Falta información sobre la duración para cuatro usuarias en el pasado.
 (Los intervalos de confianza al 95% y el test de χ^2 para una tendencia lineal en las tasas se calcularon utilizando las fórmulas indicadas en el Anexo 6.1.)

Tabla 8.5.

Incidencia de cáncer de mama en enfermeras de 45-49 años en el momento de su inclusión en la cohorte, según la duración del uso de anticonceptivos orales^a.

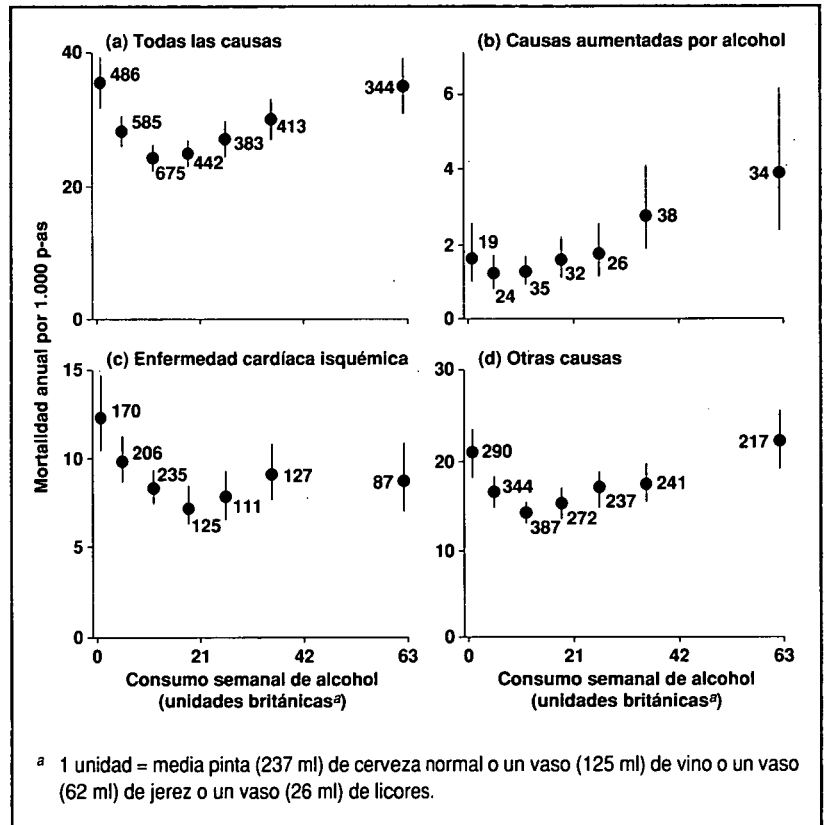
En el Ejemplo 8.13, las no usuarias de anticonceptivos orales se consideraron como la categoría basal de no expuestas. Las razones de tasas para cada categoría de duración de uso de anticonceptivos orales se calcularon dividiendo sus tasas respectivas por la tasa de la categoría basal. Por tanto, la razón de tasas para las usuarias actuales de anticonceptivos orales durante 48 meses o menos se calculó como 1.220 por 100.000 p-as / 187 por 100.000 p-as = 6,52 (intervalo de confianza al 95% 2,43-17,53) (Tabla 8.5). Este resultado indica que el riesgo aumenta entre las usuarias actuales a corto plazo, no obstante, esta estimación se basó en sólo cuatro casos de cáncer de mama (de ahí el amplio intervalo de confianza).

Para evaluar si existía una tendencia lineal (positiva o negativa) en las tasas al aumentar la duración del consumo, se realizó un test estadístico especial (test de χ^2 para una tendencia lineal) por separado para las usuarias actuales y las ex-usuarias (utilizando la fórmula indicada en el Apartado A6.1.3). Se observó una moderada tendencia positiva en las ex-usuarias de anticonceptivos orales pero ninguna tendencia lineal en las usuarias actuales (Tabla 8.5).

Ejemplo 8.14. En el estudio de médicos británicos descrito en los Ejemplos 8.3 y 8.10, en el cuestionario de 1978 se incluyeron preguntas adicionales sobre el consumo de alcohol. Se les preguntó sobre la frecuencia de la ingesta de bebidas alcohólicas y, si eran bebedores regulares (es decir, si bebían la mayoría de las semanas), el promedio de ingesta alcohólica semanal. Hacia 1991 casi un tercio de los 12.321 hombres que contestaron había fallecido. El riesgo de fallecer se examinó en relación al consumo de alcohol indicado (Doll et al., 1994b). Algunos de los resultados se presentan en la Figura 8.2.

Figura 8.2.

Mortalidad por causas específicas según el consumo de alcohol semanal: a) todas las causas; b) enfermedades relacionadas con el alcohol (p.e. cánceres de hígado, boca, faringe, esófago y laringe, cirrosis, alcoholismo y causas externas); c) enfermedad cardíaca isquémica; d) otras causas conocidas (enfermedades vasculares cerebrales, enfermedades respiratorias, todos los demás cánceres no incluidos en b), y otras). Los puntos y las barras son tasas e intervalos de confianza al 95% ajustados según edad, hábito de fumar e historia de enfermedad previas, los valores son números de muertes (reproducido con la autorización del BMJ Publishing Group, de Doll et al., 1994b).



Cabe destacar que una relación dosis-respuesta no tiene que ser necesariamente lineal. Esto se ilustra claramente en la Figura 8.2. Por ejemplo, la relación entre consumo de alcohol y la mortalidad global (Figura 8.2a) tiene

forma de U, mostrando la mortalidad más baja los hombres que declararon beber 8-14 unidades de alcohol a la semana. Por el contrario, la relación entre el consumo de alcohol y la mortalidad por causas relacionadas con el alcohol (Figura 8.2b) es básicamente lineal, con un aumento progresivo en la mortalidad al aumentar la ingesta. Por tanto, la falta de tendencia lineal (evaluada por el test de χ^2 para la tendencia) no implica la ausencia de relación. La identificación de la forma de una relación dosis-respuesta debe realizarse principalmente por representación gráfica de los datos como en la Figura 8.2. Si la forma indica una tendencia no lineal, para evaluar su significación estadística deben usarse unos métodos estadísticos especiales que están fuera del objetivo de este libro.

Debido a que en los estudios de cohorte la asignación de los miembros del estudio a las diferentes categorías de exposición no es aleatoria, es probable que las categorías de exposición difieran en muchos otros aspectos aparte de la exposición de interés. Estas diferencias deben tenerse en cuenta en el análisis. Por ejemplo, la edad es un factor de confusión importante de la relación entre el consumo de anticonceptivos orales y el cáncer de mama, ya que está fuertemente asociado con el uso de anticonceptivos orales y es de por sí un factor de riesgo independiente para el cáncer de mama. Por tanto, las diferencias en la distribución de edad entre mujeres de categorías diferentes de uso de anticonceptivos orales pueden distorsionar la relación entre éste y la incidencia de cáncer de mama. Para reducir este efecto potencial de confusión, limitamos intencionadamente el análisis de los Ejemplos 8.12 y 8.13 a un estrato de edad estrecho (mujeres de 45-49 años en el momento de su inclusión en la cohorte). No obstante, es posible (y deseable) obtener medidas (para todas las edades combinadas) 'ajustadas' por la edad y cualquier otra variable potencial de confusión utilizando métodos estadísticos más complejos que los descritos hasta ahora en este apartado. La estandarización es uno de estos métodos como veremos más adelante. La interpretación de estas 'medidas ajustadas' es similar, independientemente del método utilizado. En el Ejemplo 8.14 (y Figura 8.2), se ajustaron las tasas por la edad y el hábito de fumar. Esto significa que las diferencias de la mortalidad observadas entre las diferentes categorías de consumo de alcohol no pueden explicarse por las diferencias de edad o de hábito tabáquico, siempre que las determinaciones de estas variables de confusión fueran válidas. Estos aspectos se comentarán con más detalle en los Capítulos 13 y 14.

Otro método habitual de presentar los resultados de los estudios de cohorte, especialmente para aquellos que utilizan datos de la población general como grupo de comparación, es por el cálculo de la *Razón de mortalidad (o de incidencia) estandarizada* (RMS) (véase Apartado 4.3.3). Imagine que se observaron un total de 24 muertes por cáncer de pulmón en una cohorte de 17.800 hombres trabajadores en la industria del aislamiento mediante amianto. Este número observado (O) se compara a continuación con el número de fallecidos que se habrían esperado (E) si la cohorte tuviera las mismas tasas de mortalidad específicas por cáncer de pulmón y por edad que

toda la población masculina residente en la misma zona. Utilizando cálculos similares a los mostrados en el Apartado 4.3.3, sólo se habrían esperado siete muertes. Por tanto, la RMS es igual a $O/E = 24 / 7 = 3,4$ (ó 340 si la RMS se expresa como porcentaje). De hecho, esta medida es una *razón de tasas ajustada por la edad*. En este ejemplo, los aisladores que utilizaban amianto tenían una probabilidad 3,4 veces mayor de morir por cáncer de pulmón que el conjunto de la población masculina residente en la misma zona, y esta diferencia en la mortalidad no era debida a diferencias en la estructura de la edad entre la cohorte y la población general. Aunque este método se utiliza con frecuencia para ajustar por edad también puede utilizarse para ajustar por cualquier otra variable de confusión (p.e., tiempo de calendario, hábito de fumar). Los tests estadísticos y los intervalos de confianza al 95% para una RMS pueden calcularse como se muestra en los Apartados A6.1.2. y A6.1.3.

Ejemplo 8.15. En la cohorte ocupacional descrita en el Ejemplo 8.7, la mortalidad de 6.678 hombres trabajadores del caucho se comparó con la de la población masculina de los Estados Unidos en 1968 (McMichael et al., 1974). En la Tabla 8.6. se muestra la mortalidad por causas específicas.

Tabla 8.6.
Mortalidad por causas específicas en una cohorte de varones trabajadores del caucho.^a

Causa de muerte (código CIE-8)	Muertes observadas (O)	Muertes esperadas (E) ^b	RMS (%) ($100 \times O/E$) ^c	Intervalo de confianza 95%
Todas las causas	489	524,9	93	85-102
Todas las neoplasias (140-239)	110	108,9	101	83-122
Todas las neoplasias malignas (140-209)	108	107,3	100	82-120

^a Datos de McMichael et al., 1974).

^b Muertes esperadas calculadas basándose en las tasas específicas por edad de varones de los EE.UU., en 1968.

^c $P > 0.10$ para todas las RMS mostradas en la tabla.

(Los tests estadísticos e intervalos de confianza se calcularon utilizando las fórmulas presentadas en el Anexo 6.1.)

Otra forma de analizar los datos de cohorte, que también tiene en cuenta las diferentes duraciones del seguimiento, es utilizar los métodos de *análisis de supervivencia*. Éstos se comentarán en el Capítulo 12.

8.8 Estudios de cohorte con estudios caso-control anidados

En un estudio de cohorte tradicional, todos los individuos del estudio se someten a los mismos procedimientos –entrevistas, exámenes médicos, de-

terminaciones de laboratorio, etc. –en el momento de su inclusión en el estudio y durante el periodo de seguimiento. De manera alternativa, una cohorte puede identificarse y seguirse hasta que se desarrolle un número suficiente de casos. Entonces se obtiene y analiza información más detallada, pero sólo para los ‘casos’ y para una muestra de los individuos no enfermos (‘controles’); no para todos los miembros de la cohorte. Este tipo de estudio caso-control realizado dentro de una cohorte fija se denomina un *estudio caso-control anidado* (véase Capítulo 9). Este método es especialmente útil al tratar con procedimientos complejos y caros que de otro modo tendrían que realizarse en un número muy elevado de individuos. Por ejemplo, las muestras de sangre para todos los miembros de la cohorte pueden tomarse y congelarse en el momento de la inclusión en el estudio. No obstante, sólo se analizan, al final del seguimiento, las muestras de sangre de los casos, es decir, de aquellos individuos en la cohorte que adquieren la enfermedad en estudio y las muestras de un subgrupo de individuos que permanecen libres de enfermedad (controles).

Ejemplo 8.16. *En 1972, se estableció una cohorte de 42.000 niños en el Distrito del Oeste del Nilo en Uganda para investigar el papel etiológico del virus Epstein-Barr (VEB) en el linfoma de Burkitt (LB). En el momento de su inclusión en el estudio se obtuvo una muestra de sangre de estos niños. Al final del seguimiento en 1979, se habían detectado 16 casos nuevos de LB. El nivel de anticuerpos contra VEB en cada uno de los sueros recogidos ‘anteriormente’ a la aparición del LB en los casos se comparó entonces con los niveles de anticuerpos en los sueros de cuatro o cinco niños –de la misma edad, sexo y vecindario– de los que se había obtenido una muestra de sangre al mismo tiempo que los casos de LB pero que no habían desarrollado un LB (‘controles’) (Geser et al., 1982).*

En el Ejemplo 8.16, se obtuvieron y conservaron muestras de sangre de los 42.000 niños que participaron en el estudio, pero las pruebas víricas bastante complejas y costosas se efectuaron sólo en las muestras de suero de los 16 niños que desarrollaron el linfoma y en una muestra de aproximadamente 80 niños no enfermos que actuaron como controles.

Del mismo modo, en estudios de cohorte nutricionales pueden utilizarse diarios de alimentos para medir la ingesta habitual en la dieta del sujeto. No obstante, la codificación y análisis de los diarios de alimentos es un trabajo intenso, muy laborioso y caro. Por esta razón, en lugar de analizar los diarios de cada miembro de la cohorte, puede realizarse un estudio caso-control anidado en el que sólo se examinen los diarios de los casos y de una muestra de miembros de la cohorte sin la enfermedad de interés (‘controles’).

Este tipo de diseño de estudio y el análisis de sus datos se comentarán más ampliamente en los Apartados 9.3.3 y 9.5.

8.9 Interpretación

La principal ventaja de los estudios de cohorte es que la *secuencia temporal* entre la exposición a un factor de riesgo y la aparición de la enfermedad se observa claramente.

No obstante, los estudios de cohorte pueden sufrir sesgos importantes. El conocimiento de la presencia o ausencia de la exposición a un factor de riesgo en particular puede afectar al subsiguiente diagnóstico de la enfermedad e introducir un *sesgo de determinación*. Esto puede producirse si la decisión sobre la presencia o ausencia de enfermedad la efectúan personas que conocen el estado del sujeto respecto al factor en estudio. Los estudios de cohorte también son potencialmente susceptibles de un *sesgo de selección* debido a la pérdida de sujetos en estudio. Tales pérdidas pueden producirse inicialmente si una parte de la población en estudio no participa, o más adelante en el curso del estudio cuando una parte de los participantes se pierde en el seguimiento. Estas pérdidas no invalidan necesariamente el estudio. No obstante, los investigadores deben valorar si las razones de la pérdida pueden haber afectado considerablemente a los resultados del estudio. A veces se puede conseguir información adicional sobre los sujetos perdidos, especialmente si abandonaron el estudio debido a enfermedad o muerte por razones que puedan estar relacionadas con las exposiciones y los resultados investigados.

Al igual que con cualquier otro estudio observacional, la *confusión* es un aspecto crítico en la interpretación de los estudios de cohorte. Pueden utilizarse técnicas estadísticas especiales para tener en cuenta el efecto de las variables potenciales de confusión, pero sólo si estas variables se conocían en el momento de la obtención de los datos y si se midieron correctamente. Si no se obtuvieron estos datos, deberá evaluarse en qué medida los resultados observados pueden estar afectados por confusión basándose en todas las evidencias biológicas y epidemiológicas disponibles.

En el Ejemplo 8.17, los mineros de uranio presentan una mortalidad significativamente mayor para los tres tipos de cáncer respecto a la población general (Tabla 8.7). Antes de concluir que estos riesgos aumentados eran debidos a exposiciones en las minas, hay que considerar las posibles explicaciones alternativas para los resultados observados. Una posible explicación para el riesgo aumentado de cáncer de pulmón es que los mineros fumaban más que la población general. No se disponía de información sobre el hábito de fumar de los mineros, pero un estudio en 697 hombres checos mineros de uranio en 1974 mostró que un 76% eran fumadores, porcentaje más alto que la media masculina en la antigua Checoslovaquia a principios de los años 70 (66%) (véase una discusión completa en Tomášek *et al.* (1994)). Los resultados de este estudio indican que es poco probable que las diferencias halladas en el hábito de fumar justifiquen un riesgo 5 veces mayor de morir por cáncer de pulmón entre los mineros respecto a la población general. Además, la Tabla 8.8 (Ejemplo 8.8) muestra una tendencia positiva significativa entre la mortalidad por cáncer de pulmón y la exposición acumulada creciente a radón. Esta tendencia proporciona una sólida

justificación de la existencia de una asociación real entre la exposición a radón y el cáncer de pulmón. Esta conclusión se refuerza por haberse observado resultados similares en otros mineros expuestos a radón.

Ejemplo 8.17. En 1970 se identificó una cohorte de mineros de uranio en el oeste de Bohemia mediante una búsqueda retrospectiva de los registros ocupacionales. Los trabajadores se incluyeron en el estudio si: a) empezaron a trabajar bajo tierra en el periodo 1948-59; b) permanecieron en este trabajo como mínimo cuatro años; c) sus registros personales y de empleo estuvieran disponibles. Un total de 4.320 mineros se incluyeron en el estudio. Los mineros estuvieron expuestos a niveles altos de radón y de polvo. En una de las dos minas principales, también estuvieron expuestos a arsénico. La mortalidad de estos mineros hasta finales de 1990 (una media de 25 años de seguimiento) se comparó con la mortalidad de la población general de la antigua Checoslovaquia. No se disponía de información sobre el consumo de alcohol y tabaco de los mineros (Tomášek et al., 1993, 1994). La mortalidad por cáncer en esta cohorte según algunas localizaciones tumorales se muestra en la Tabla 8.7.

Causa de muerte	Muertes observadas (O)	Muertes esperadas (E) ^b	RMS (O/E) ^b (intervalos de confianza 95%)
Cáncer de pulmón	704	138,6	5,08 (4,71-5,47)
Cáncer de hígado	22	13,2	1,67 (1,04-2,52)
Cáncer de vesícula biliar y conductos biliares extrahepáticos	12	5,3	2,26 (1,16-3,94)

^a Datos de Tomášek et al. (1993).

^b Número de muertes esperadas (E) calculadas utilizando las tasas nacionales de mortalidad específicas por edad en la antigua Checoslovaquia.

Tabla 8.7.

Número de muertes observadas (O) por localizaciones tumorales seleccionadas en la cohorte de mineros de uranio en el oeste de Bohemia, en comparación con el número esperado (E) si los mineros presentaran la misma mortalidad que la población masculina general de la antigua Checoslovaquia^a.

También se observó una tendencia positiva en la mortalidad por cáncer de vesícula biliar y conductos biliares extrahepáticos con niveles crecientes de exposición acumulada al radón, pero las evidencias a favor de esta observación en otros estudios epidemiológicos son escasas y son necesarias más investigaciones para confirmarla.

Por el contrario, la mortalidad por cáncer de hígado no aumentó con la exposición acumulada a radón, haciendo poco probable que el exceso de casos observado en los mineros estuviera causado por el radón. No se disponía de información sobre el consumo de alcohol de los mineros, pero estaban bien pagados en comparación con otros trabajadores checos, y por tanto, es probable que su consumo de alcohol fuera mayor que en la población general. También presentaron un exceso significativo de muertes por cirrosis hepática, probablemente causada por el consumo de alcohol y, en seis de las muertes por cáncer de hígado, también se mencionó la cirrosis en el certificado de defunción (Tomášek et al., 1993).

Ejemplo 8.18. En la cohorte de mineros de uranio ilustrada en el Ejemplo 8.17, se realizó una estimación de la exposición acumulada al gas radón (y sus derivados) para cada uno de los mineros. La exposición se determinó en unidades de 'meses de trabajo en cada nivel de exposición' [Working Level Months (WLM)], que se calculó considerando el tiempo consumido en cada mina en conjunción con aproximadamente 39.000 determinaciones de gas radón en pozos específicos efectuadas en 1949-63 (Tomášek et al., 1993). La mortalidad en función del nivel de exposición se muestra en la Tabla 8.8.

Tabla 8.8. Número de muertes observadas (O) y razón de mortalidad estandarizada (O/E) según localización tumoral y la exposición acumulada a radón^a.

Localización tumoral		Exposición acumulada a radón (WLM)					Valor de p para tendencia
		<110	110-149	150-209	210-329	≥330	
Cáncer de pulmón	O	86	100	139	161	181	<0,001
	O/E ^b	3,07	3,66	4,98	6,23	8,10	
Cáncer de hígado	O	7	3	4	5	3	0,57
	O/E ^b	2,70	1,18	1,53	2,04	1,40	
Cáncer de vesícula biliar y conductos biliares extrahepáticas	O	0	2	1	3	6	0,003
	O/E ^b	0,00	1,92	0,93	3,03	6,73	

^a Datos de Tomášek et al. (1993)

^b Número esperado de muertes (E) calculado utilizando las tasas nacionales de mortalidad masculina en la antigua Checoslovaquia.

Los aspectos que requieren ser tratados en la interpretación de los resultados de los estudios de cohorte se comentarán más ampliamente en el Capítulo 13.

Cuadro 8.1. Puntos esenciales

- Los estudios de cohorte son estudios en los que los sujetos se eligen basándose en su exposición y se siguen en el tiempo. No obstante, a diferencia de los estudios de intervención, la asignación de la exposición no la determinan los investigadores.
- Las principales ventajas de este tipo de estudio son:
 1. La exposición se determina antes del inicio de la enfermedad y por tanto es probable que no esté sesgada en términos de desarrollo de la enfermedad.
 2. Las exposiciones poco frecuentes pueden investigarse mediante la selección apropiada de las cohortes de estudio.
 3. Se pueden estudiar múltiples resultados finales (enfermedades) para cada una de las exposiciones.
 4. La incidencia de la enfermedad puede determinarse para los grupos de expuestos y no expuestos.
- Los principales inconvenientes de este tipo de estudio son:
 1. Pueden ser muy caros y requerir mucho tiempo, especialmente si se realiza de manera prospectiva.
 2. Los cambios de la exposición en el tiempo y en los criterios de diagnóstico pueden afectar a la clasificación de los individuos.
 3. La identificación de la enfermedad puede estar influenciada por el conocimiento del estado de exposición del sujeto (*sesgos de información*).
 4. Las pérdidas del seguimiento pueden introducir *sesgos de selección*.
- Los estudios de cohorte en epidemiología del cáncer generalmente implican el seguimiento de un gran número de individuos durante periodos prolongados de tiempo. Por tanto, tienden a ser muy caros y requerir mucho tiempo. Pueden utilizarse diferentes estrategias para reducir sus costes:
 1. Utilizar registros (o muestras biológicas) preexistentes para identificar retrospectivamente una población de estudio adecuada y obtener información sobre el estado de exposición de sus miembros (estudio de cohorte histórico).
 2. Utilizar sistemas de vigilancia disponibles (p.e., certificados de defun-

Lecturas adicionales

* El libro de Breslow y Day (1987) proporciona una explicación muy completa del papel, diseño, análisis e interpretación de los estudios de cohortes en epidemiología del cáncer. Parte del material de este libro se presenta a un nivel relativamente avanzado.

* En los artículos de Carpenter (1987) y McMichael (1976) puede encontrarse una discusión sobre el efecto del trabajador sano.

ción, registros de cáncer) para seguir a los sujetos y obtener información sobre las enfermedades de interés.

3. Utilizar tasas nacionales (o locales) como grupo de comparación no expuesto.
4. Realizar un estudio caso-control anidado.

Capítulo 9

Estudios caso-control

Un estudio caso-control implica la identificación de individuos con ('casos') y sin ('controles') una enfermedad o condición particular. La prevalencia (o nivel) de exposición a un factor se determina entonces en cada grupo y se compara. Si la prevalencia de exposición entre casos y controles es diferente, entonces puede inferirse que la exposición puede estar asociada a una aparición aumentada o disminuida del resultado de interés (véase Apartado 9.5).

***Ejemplo 9.1.** La relación entre el uso de estrógenos conjugados y el riesgo de cáncer de endometrio se examinó en 188 mujeres blancas de 40-80 años con cáncer de endometrio recién diagnosticado y 428 controles de edad similar hospitalizados por enfermedades no cancerosas que requerían cirugía. El estudio se realizó en el Hospital para Mujeres Parkway Division, en Boston, Massachusetts, entre Enero de 1970 y Junio de 1975. Los datos sobre el uso de fármacos y la historia reproductiva se obtuvieron de las historias médicas del hospital y de las historias clínicas de los médicos particulares de cada mujer. Un 39% de los casos y un 20% de los controles tenía historia de consumo previo de estrógenos conjugados (Buring et al., 1986).*

En el Ejemplo 9.1, se identificaron mujeres con ('casos') y sin ('controles') cáncer de endometrio y se obtuvo información sobre el consumo en el pasado de estrógenos conjugados ('exposición') a partir de registros médicos del hospital y otros. La prevalencia de uso de estrógenos conjugados fue mucho mayor entre los casos (39%) que entre los controles (20%) sugiriendo que el uso de este fármaco estaba asociado a un aumento de la incidencia de cáncer de endometrio.

La diferencia principal entre los estudios de cohorte y caso-control reside en la selección de los sujetos a estudio. En un estudio de cohorte, se empieza seleccionando a los sujetos que inicialmente no están enfermos y se clasifican según su exposición a factores de riesgo putativos (Capítulo 8), mientras que en un estudio caso-control se empieza por identificar a los sujetos basándose en la presencia o ausencia de la enfermedad (o cualquier otro resultado) a estudio y se determina la exposición previa a los factores de riesgo putativos.

Los estudios caso-control son especialmente adecuados para estudiar enfermedades relativamente raras con un periodo de inducción prolongado (p.e. cáncer). Esto es debido a que un estudio caso-control empieza con su-

jetos que ya han desarrollado la enfermedad de interés, de modo que no es necesario esperar a que pase el tiempo entre la exposición y la aparición de la enfermedad como en los estudios de cohorte prospectivos. Los estudios de cohorte históricos permiten un ahorro de tiempo similar pero sólo pueden realizarse en las situaciones raras en las que se han conservado registros antiguos con datos pertinentes sobre exposiciones o cuando los bancos de muestras biológicas se han conservado correctamente y se dispone de unas pruebas de laboratorio adecuadas para la determinación de las exposiciones de interés.

9.1 Hipótesis del estudio

Al igual que con cualquier otro tipo de estudio, la hipótesis específica en estudio debe establecerse claramente antes de considerar los detalles del diseño del estudio caso-control. Si no se realiza esto, puede dar lugar a un diseño pobre y a problemas en la interpretación de los resultados. Los estudios caso-control permiten la evaluación de exposiciones múltiples que pueden estar relacionadas con una enfermedad específica (así como posibles interacciones entre ellas). El Ejemplo 9.2 ilustra claramente este punto.

***Ejemplo 9.2.** En España y Colombia se realizó un estudio caso-control basado en la población para evaluar la relación entre el cáncer de cuello uterino y la exposición al virus del papiloma humano (VPH), aspectos seleccionados de comportamiento sexual y reproductor, uso de anticonceptivos orales, prácticas de cribaje, hábito de fumar y posibles interacciones entre ellos. El estudio incluyó 436 casos incidentes de carcinoma del cuello del útero de células escamosas invasivo confirmado por histología y 387 controles de edades similares elegidos aleatoriamente de la población general que generó los casos (Muñoz, et al., 1992a).*

Los estudios caso-control constituyen con frecuencia una de las primeras aproximaciones al estudio de la etiología de una enfermedad o condición en particular. Esto es debido, en parte, al igual que en el Ejemplo 9.3, a la posibilidad de estudiar muchas exposiciones y en parte porque pueden realizarse de manera relativamente barata y rápida.

***Ejemplo 9.3.** Debido a su baja frecuencia, se conoce muy poco sobre la etiología de los tumores malignos de células germinales (TMCG) en niños. Con objeto de estudiar los factores de riesgo de estas enfermedades y generar hipótesis etiológicas, se realizó un estudio caso-control basado en la población con 105 niños con TMCG y 639 controles (Shu et al., 1995).*

Los resultados de estos estudios caso-control *exploratorios* pueden indicar hipótesis específicas que pueden ser verificadas subsiguientemente en estudios diseñados específicamente para ello.

9.2 Definición y selección de los casos

9.2.1 Definición de caso

Para la definición de un caso es esencial tener unos criterios precisos. Normalmente, se recomienda solicitar documentación objetiva de que los casos realmente padecen la enfermedad o condición de interés, incluso a costa de eliminar algunos casos reales. Por ejemplo, para la mayoría de los cánceres debe solicitarse un diagnóstico confirmado por histología. Al aceptar casos peor documentados, el investigador corre el riesgo de diluir el grupo de casos con algunos que no sean casos, disminuyendo las oportunidades de encontrar una diferencia real entre casos y controles en la prevalencia a la exposición.

A veces es imposible eliminar todos los casos cuyo diagnóstico no está documentado correctamente, especialmente si el conjunto de casos disponible es relativamente pequeño. En estas circunstancias, pueden clasificarse los casos por la certeza del diagnóstico. Esta clasificación proporciona una oportunidad de evaluar la probabilidad de que los resultados estén afectados por una clasificación errónea de la enfermedad (Capítulo 13). Por ejemplo, los casos en un estudio caso-control particular pueden clasificarse como de diagnóstico 'seguro', 'probable', y 'posible'. Si existe una clasificación errónea de la enfermedad, en el análisis se observará una disminución gradual del riesgo relativo desde la categoría 'seguro' a la categoría 'posible', ya que la probabilidad de que los que no son casos puedan haberse diagnosticado erróneamente como casos aumenta de la categoría 'seguro' a la categoría 'posible'.

La definición de caso debe establecerse de tal modo que no existan ambigüedades sobre los tipos de casos y las fases de la enfermedad a incluir o excluir del estudio. La elección de los casos debe guiarse más por la preocupación de validez que por la posibilidad de generalizar. Por ejemplo, en un estudio de cáncer de mama puede aprenderse más limitando los casos (y los controles) a las mujeres pre- o post-menopáusicas que incluyendo mujeres de todas las edades (a no ser que el número de casos en cada grupo sea lo suficientemente grande como para permitir análisis separados) ya que los factores de riesgo para el cáncer de mama pre- y post-menopáusico pueden ser diferentes. Asegurando que los casos sean un grupo relativamente homogéneo, se aumentan las oportunidades de detectar relaciones etiológicas importantes. La posibilidad de generalizar los resultados a una población entera normalmente es menos importante que establecer una relación etiológica, incluso cuando sólo sea para un subgrupo pequeño de la población.

En la selección de los casos deben incluirse únicamente aquellos para los cuales existe una posibilidad razonable de que la enfermedad haya sido producida por la exposición a estudio.

En el Ejemplo 9.4, los casos se limitaron a mujeres nacidas a partir de los años 20 porque sólo podían haber estado expuestas al factor de interés (es decir, anticonceptivos orales) las mujeres nacidas después de esta fecha.

***Ejemplo 9.4.** Se realizó un estudio caso-control, basado en hospitales, para evaluar la relación entre el uso de anticonceptivos orales y el riesgo de desarrollar alguno de seis cánceres específicos diferentes. El estudio se realizó en 10 centros participantes de ocho países (Chile, China, Colombia, Israel, Kenya, Nigeria, Filipinas y Tailandia) desde octubre de 1979 a septiembre de 1986. Las mujeres con cánceres recién diagnosticados de mama, cuerpo del útero, cuello del útero, ovario, hígado y ductus biliar eran elegibles si habían nacido después de 1924 o de 1929 (dependiendo de cuándo estaban disponibles localmente los anticonceptivos orales) y habían vivido en la zona del hospital participante durante un año como mínimo (The WHO Collaborative Study of Neoplasia and Steroid Contraceptives, 1989).*

Aunque la mayoría de los estudios caso-control sólo incluyen un grupo de casos, pueden estudiarse simultáneamente dos o más cánceres cuyos factores de riesgo se cree que comparten los mismos factores de riesgo o relacionados. El Ejemplo 9.4 ilustra este punto. Estos estudios caso-control de múltiples enfermedades pueden considerarse como una serie de estudios caso-control. Esta estrategia proporciona dos ventajas principales. En primer lugar, proporciona la oportunidad de estudiar más de un cáncer por un coste adicional relativamente bajo. En segundo lugar, los grupos control pueden combinarse para proporcionar a cada comparación caso-control un mayor poder estadístico, es decir, la capacidad del estudio de detectar un efecto real, si realmente existe uno, se potencia debido al mayor número de controles por caso (véase Capítulo 15).

Si la enfermedad o condición de interés es muy rara, el estudio puede requerir varios centros participantes posiblemente situados en diferentes países. El estudio presentado en el Ejemplo 9.4 se realizó en 10 centros en ocho países. A pesar de esto, sólo se acumularon 122 casos diagnosticados de cáncer de hígado (el más raro de todos los tipos de cáncer examinados) durante el periodo de estudio de siete años. Algunos estudios incluyen intencionadamente centros en zonas de incidencia baja y alta para evaluar si los factores de riesgo son similares. Por ejemplo, el estudio de cáncer de cuello uterino mencionado en el Ejemplo 9.2 se realizó en Colombia y España, países con una diferencia de 8 veces en la incidencia de cáncer de cuello uterino (Muñoz *et al.*, 1992a).

***Ejemplo 9.5.** En el estudio caso-control de cáncer de cuello uterino mencionado en el Ejemplo 9.2, los casos elegibles fueron incidentes, confirmados por histología como carcinomas invasivos de cuello uterino de células escamosas e identificados entre los residentes en las zonas de estudio durante como mínimo seis meses. Se excluyeron las pacientes que por su situación física y/o mental no era recomendable la entrevista y/o toma de muestras. A su vez se excluyeron las pacientes si eran mayores de 70 años (Muñoz *et al.*, 1992a).*

Los criterios de elegibilidad no sólo deben incluir una definición clara del caso, sino también otros criterios de inclusión (Ejemplo 9.5). Deben excluirse las personas que están demasiado enfermas para cooperar o para las cuales los métodos del estudio puedan ocasionar un trastorno físico o psicológico considerable. También se suele excluir a personas mayores en estudios caso-control de cáncer, debido a que es probable que su diagnóstico sea menos válido y por su dificultad en recordar las exposiciones en el pasado.

Normalmente, no es posible la inclusión de todos los pacientes que cumplen los criterios de selección por una serie de razones. Los sujetos pueden cambiar de zona, morir o, simplemente, negarse a cooperar. El investigador debe indicar cuántos casos cumplían los criterios de inclusión iniciales, las razones de cualquier exclusión, y el número omitido por cada razón (como en el Ejemplo 9.6). Esta información permite evaluar el grado en que los resultados del estudio pueden estar afectados por un sesgo de selección (véase Capítulo 13).

Ejemplo 9.6. Para evaluar el papel de los factores dietéticos en la etiología de cáncer de estómago y su contribución a la variación geográfica importante en la mortalidad por este cáncer en el país, se realizó en Italia un gran estudio caso-control multicéntrico en zonas de alto y bajo riesgo. Eran elegibles como casos todos los pacientes con cáncer de estómago confirmado por histología diagnosticados por primera vez entre junio de 1985 y diciembre de 1987, residentes en las zonas de estudio y de 75 años o menores. Se identificaron un total de 1.129 casos en los departamentos de cirugía y gastroenterología y en los servicios ambulatorios de gastroscopia de hospitales privados y públicos. Aproximadamente un 83% de estos casos se entrevistaron utilizando un cuestionario estructurado (Buiatti et al., 1989a,b). La Tabla 9.1 muestra el número de pacientes elegibles en cada centro de reclutamiento, cuántos se incluyeron y las razones de exclusión.

	Elegibles n (%)	Incluidos n (%)	Excluidos por		
			Rechazo n (%)	Mala salud n (%)	Fallecidos ^b n (%)
Cagliari	104 (100)	82 (78,9)	3 (2,9)	4 (3,8)	15 (14,4)
Cremona	71 (100)	66 (93,0)	0 (0,0)	4 (5,6)	1 (1,4)
Florenia	435 (100)	382 (87,8)	9 (2,1)	28 (6,4)	16 (3,7)
Forlì	255 (100)	232 (91,0)	8 (3,1)	14 (5,5)	1 (0,4)
Génova	155 (100)	122 (78,7)	3 (1,9)	24 (15,5)	6 (3,9)
Imola	76 (100)	47 (61,8)	9 (11,8)	8 (10,5)	12 (15,9)
Siena	133 (100)	85 (63,9)	18 (13,5)	29 (21,8)	1 (0,8)
Total	1.229 (100)	1.016 (82,7)	50 (4,1)	111 (9,0)	52 (4,2)

^a Datos de Buiatti et al. (1989a).

^b Fallecidos después de ser identificados como casos potenciales.

Tabla 9.1.

Niveles de participación entre los casos de cáncer de estómago y razones para la no participación por centro de reclutamiento^a.

Siempre que sea posible, debe buscarse información sobre toda la serie de casos elegibles, considerando características tales como sexo, educación, situación socioeconómica, etc., de modo que puedan evaluarse los factores de selección para los sujetos que no participan. Esta información puede obtenerse a partir de archivos hospitalarios y de registros de cáncer (como en el Ejemplo 9.7).

Ejemplo 9.7. En el estudio caso-control de cáncer de estómago comentado en el Ejemplo 9.6, se compararon el número y características de los casos incluidos en cada centro participante con la información obtenida de los registros locales de cáncer o los departamentos de patología. La Tabla 9.2 muestra que los casos incluidos en el estudio en Florencia eran algo más jóvenes y con mayor frecuencia mujeres que los casos notificados en el registro del cáncer local. Esto era debido a que se habían excluido del estudio los casos sin confirmación, y éstos generalmente eran hombres en los grupos de mayor edad (Buiatti *et al.*, 1989a, b).

Tabla 9.2.

Distribución (%) por edad y sexo de los casos de cáncer de estómago incluidos en el centro de Florencia durante 1985-87 y de los casos de cáncer de estómago notificados al registro local de cáncer en 1985^a.

Edad (años)	Hombres (H)				Mujeres (M)				H:M ratio
	<45	45-54	55-64	65-74	<45	45-54	55-64	65-74	
Casos incluidos, 1985-87	4,4	13,3	33,2	49,1	3,8	12,1	18,2	65,9	1,7
Casos notificados al registro, 1985	2,6	10,7	33,2	53,6	3,1	13,5	20,8	62,5	2,0

^a Datos de Buiatti *et al.* (1989a,b). (En estos artículos no se indicó el número de casos en que se basaban estos porcentajes).

9.2.2 Casos incidentes o casos prevalentes

Un aspecto importante a considerar en la etapa de diseño de un estudio caso-control es si incluir los casos *prevalentes* o sólo los *incidentes*. Los casos incidentes son todos los casos *nuevos* que aparecen en una población dentro de un periodo de tiempo establecido. Los casos prevalentes son todos los casos *existentes* (nuevos y anteriores) presentes en una población en un momento en el tiempo (o dentro de un periodo de tiempo muy corto) (véase Apartado 4.2). El principal inconveniente de utilizar una serie de casos prevalentes es que los pacientes con una duración prolongada de la enfermedad tienden a estar sobrerrepresentados debido a que, por definición, todos aquellos con una corta duración desaparecen del conjunto de casos prevalentes por recuperación o muerte. A no ser que pueda justificarse la suposición de que la exposición estudiada no está relacionada con la recuperación o supervivencia, debe hacerse todo lo posible para limitar la inclusión a los casos incidentes. El estudio caso-control, al igual que un estudio de cohorte, tiene como objetivo identificar los factores responsables del desarrollo de la enfermedad utilizando únicamente los casos recién diagnosticados (incidentes) y seleccionando controles que sean representativos de la población de procedencia de los casos. Además, los casos prevalentes pueden no ser representativos de todos los casos, cuando algunos pacientes

afectados se institucionalizan en otro lugar o se trasladan a otra ciudad en la que existan instalaciones especiales para el tratamiento.

También existen otras ventajas del uso de casos incidentes. La memoria de acontecimientos pasados en historias personales tiende a ser más exacta en los casos recién diagnosticados que en los casos prevalentes. Además, es menos probable que los casos incidentes hayan cambiado sus costumbres (o 'exposiciones') como resultado de la enfermedad.

Si la utilización de casos prevalentes se hace inevitable por limitaciones de tiempo o recursos, los casos deben incluirse inmediatamente (o lo más próximos posible) después del diagnóstico. La comparación de las frecuencias (o niveles) de exposición entre los sujetos diagnosticados en diferentes momentos permite comprobar las características de los casos prevalentes. Si la frecuencia de exposición a un factor de riesgo para la enfermedad es distinta entre los casos según el intervalo transcurrido desde el momento del diagnóstico, debe sospecharse la presencia de un *sesgo de supervivencia*. Por ejemplo, si los casos expuestos al factor de estudio presentan una supervivencia menor que los no expuestos, en una serie de casos prevalentes los casos expuestos estarán menos representados a medida que aumente el tiempo desde el diagnóstico. Como resultado, disminuirá la prevalencia de la exposición registrada entre los casos supervivientes.

A veces, cuando es difícil establecer una fecha específica de inicio de la enfermedad, no es posible excluir los casos prevalentes. Por ejemplo, los estudios caso-control para examinar factores de riesgo de infección por *Helicobacter pylori* deben basarse en los casos prevalentes debido a que es difícil establecer la fecha de inicio de esta infección.

9.2.3 Fuentes de casos

En un estudio debe considerarse detenidamente qué casos deben ser incluidos. El estudio puede estar '*basado en un hospital*' en cuyo caso se incluyen todos los pacientes que cumplen los criterios de elegibilidad y acuden a un determinado hospital o un grupo de hospitales. En el Ejemplo 9.1, los casos eran mujeres blancas, de 40-80 años, ingresadas en un hospital de Boston desde enero de 1970 a junio de 1975 con un primer diagnóstico de cáncer de endometrio.

Como alternativa, el estudio puede estar '*basado en la población*' y los casos se toman de una población definida durante un periodo de tiempo definido. Esto se muestra en el Ejemplo 9.8.

En estudios caso-control basados en la población es fundamental asegurar que se han incluido la totalidad de los casos. Los aspectos que deben considerarse son la accesibilidad de los pacientes a centros sanitarios (que es probable que sea un problema secundario en estudios de cáncer en países en los que la asistencia médica está generalizada pero mucho más preocupante en países en los que no es así), la dificultad en encontrar a los sujetos y las negativas a participar por parte de algunos enfermos.

Los registros de cáncer basados en la población pueden utilizarse para incluir todos los casos incidentes de la población residente, aunque su papel

Ejemplo 9.8. En el estudio caso-control de cáncer de cuello uterino mencionado en el Ejemplo 9.2, se organizó una red activa de búsqueda de casos con visitas periódicas a todos los hospitales, clínicas y departamentos de patología en el sector público y privado en cada zona de estudio, para identificar y entrevistar los casos antes de aplicar un tratamiento. También se identificaron todos los casos de neoplasias intraepiteliales cervicales (NIC) grado III diagnosticados durante el periodo de estudio. Una serie de patólogos revisaron las muestras histológicas para asegurar la exhaustividad en la inclusión de los casos de cáncer invasivo (Muñoz et al., 1992a).

como fuente útil puede ser limitado si existe un intervalo de tiempo importante entre el diagnóstico y el registro. Los casos con una supervivencia corta pueden haber fallecido en este tiempo y otros pueden haberse trasladado de la zona de residencia como consecuencia de la enfermedad. Por tanto, en el momento en que se registran los casos puede ser difícil aceptarlos como representativos de los casos incidentes.

9.3 Definición y selección de los controles

9.3.1 Definición de los controles

Los controles deben cumplir todos los criterios de elegibilidad definidos para los casos aparte de aquellos relacionados con el diagnóstico de la enfermedad. Por ejemplo, si los casos son mujeres con cáncer de mama de 45 años o mayores, los controles deben elegirse entre mujeres sin la enfermedad en el mismo grupo de edad que los casos.

Si la enfermedad estudiada es poco frecuente en el grupo que sirve como fuente de controles, serán necesarios pocos esfuerzos de diagnóstico o documentación para descartar la presencia de enfermedad en los controles elegidos. Muchas veces será suficiente una pregunta simple en la entrevista. No obstante, si la enfermedad pudiera aparecer normalmente en los controles, es conveniente intentar reducir al máximo una clasificación errónea, como por ejemplo mediante una revisión de la historia clínica personal de cada individuo (Ejemplo 9.9).

Ejemplo 9.9. En el estudio caso-control de cáncer de cuello uterino descrito en el Ejemplo 9.2, los controles eran elegibles si eran menores de 70 años, no habían recibido tratamientos previos para el cáncer de cuello uterino o no se habían sometido a histerectomía y si el frotis citológico tomado en el momento de la inclusión era normal o presentaba únicamente cambios inflamatorios (Papanicolau grados I y II) (Muñoz et al., 1992a).

9.3.2 Fuente de controles

En estudios caso-control, los controles deben representar la población de la cual se toman los casos, es decir, deben proporcionar una estimación de la prevalencia de la exposición en la población de la cual proceden los casos. De lo contrario, es probable que los resultados del estudio estén distorsionados por un *sesgo de selección*.

En un *estudio caso-control anidado* es relativamente sencillo asegurar que los casos y los controles proceden de la misma población en estudio ya que los dos procederán de una población definida claramente –la cohorte (véase Apartado 8.8). En general, todos los casos que aparecen mientras se sigue la cohorte de manera prospectiva serán los ‘casos’ en el estudio caso-control, mientras que una muestra de los componentes de la cohorte no afectados serán identificados como ‘controles’.

Ejemplo 9.10. Se estudió la relación entre la infección por *Helicobacter pylori* y el cáncer de estómago en una cohorte de hombres japoneses americanos habitantes de Hawaii. Se incluyeron un total de 5.908 varones desde 1967 a 1970. En 1989, se habían identificado un total de 109 casos nuevos de cáncer de estómago confirmados patológicamente entre los miembros de la cohorte. En los sueros conservados de todos los pacientes con cáncer de estómago (‘casos’) y de una muestra de sujetos que no desarrollaron cáncer de estómago (‘controles’) se analizó la presencia de anticuerpos IgG séricos a *Helicobacter pylori* (Nomura et al., 1991).

En el Ejemplo 9.10, los casos y los controles se tomaron de la misma población –la cohorte de 5.908 hombres japoneses americanos habitantes de Hawaii.

Conceptualmente, puede suponerse que todos los estudios caso-control están ‘anidados’ dentro de una población particular. En un *estudio caso-control basado en la población*, una población del estudio puede definirse a partir de la población de la que se obtienen todos los casos incidentes; los controles deben elegirse aleatoriamente entre los miembros sin la enfermedad de la misma población. Por ejemplo, todos los casos nuevos de cáncer infantil diagnosticados en la zona de cobertura de un registro regional de cáncer. Los controles para estos casos se tomarían correctamente de la población de la misma zona en los mismos grupos de sexo y edad. Incluso cuando los casos se identifiquen exclusivamente en hospitales, aun puede ser razonable suponer que representan todos los casos en la zona de captura si la enfermedad es lo suficientemente grave como para que todos los casos sean hospitalizados (lo que es bastante probable para la mayoría de los casos de cáncer en países con buena asistencia sanitaria).

Por lo general, obtener controles de una muestra aleatoria de la población de captura es caro y requiere tiempo. Para realizar el muestreo es necesario disponer de una lista de todos los sujetos o familias elegibles o debe crearse una (como en el Ejemplo 9.1). (Los métodos para elegir una muestra aleato-

ria de la población en estudio se comentarán en el Capítulo 10.) Además, cabe la posibilidad de que las personas sanas tiendan a no participar, lo que puede introducir un sesgo de selección debido a la falta de colaboración.

Ejemplo 9.11. *En el estudio caso-control de cáncer de cuello uterino mencionado anteriormente, los controles se eligieron aleatoriamente de la población general que generó los casos. En Colombia, se utilizaron fotografías aéreas actualizadas de la ciudad como marco de muestreo. A partir de estas fotografías, se eligieron las casas al azar y se empleó la búsqueda puerta a puerta siguiendo una rutina predeterminada para identificar controles adecuados. En España, como marco de muestreo se utilizó el censo provincial de 1981, el último disponible (Muñoz et al., 1992a).*

Los controles también pueden elegirse entre las personas próximas al caso, tales como *amigos y parientes* de la misma población de cobertura que los casos. Aunque es relativamente fácil identificar estos controles y obtener su cooperación, existe el peligro de que sean demasiado parecidos (apareamiento excesivo) a los casos en lo que respecta a las exposiciones y otras características (véase Apartado 9.3.4). También se utilizan *controles del vecindario*, pero es probable que las personas que viven en el mismo vecindario sean similares en muchos aspectos, de modo que estos controles también pueden dar lugar a un apareamiento excesivo. Además, si el entrevistador tiene que visitar cada barrio para ponerse en contacto con estos controles, el coste del estudio puede ser extremadamente elevado.

Cuando se utilizan casos detectados en hospitales puede que no sea posible definir la población de procedencia de los casos, bien porque no puede definirse la zona exacta de atracción del hospital o bien porque no todos los casos en la zona se dirijan al hospital y los que se dirigen estén seleccionados según unos criterios particulares (p.e., la gravedad de la enfermedad). En estas circunstancias, pueden utilizarse *controles seleccionados del mismo hospital* porque la población en estudio puede definirse como 'usuarios potenciales del hospital'.

Los controles hospitalizados presentan varias ventajas. Existen muchos factores de selección que llevan a las personas a los hospitales (p.e., situación económica, zona de residencia, raza, religión) y eligiendo controles del mismo conjunto de pacientes del que proceden los casos, se reduce el efecto de estos factores. Estos controles generalmente se identifican fácilmente y tienden a cooperar. Además, debido a que también han experimentado un diagnóstico de enfermedad y una hospitalización, se asemejan a los casos en lo que respecta a su disposición a proporcionar información completa y exacta. De este modo se reducen las diferencias potenciales entre casos y controles en cuanto a la calidad del recuerdo de sus exposiciones en el pasado.

La elección de controles hospitalarios adecuados con frecuencia es difícil y debe prestarse especial atención para evitar un sesgo de selección. Un in-

conveniente importante de un grupo control elegido a partir de individuos enfermos es que algunas de sus enfermedades pueden compartir factores de riesgo con la enfermedad en estudio, es decir, pueden presentar una prevalencia de exposición mayor o menor en comparación con la población de la que proceden los casos. Por ejemplo, en un estudio sobre la relación entre alcohol y el cáncer de mama, el uso de controles del servicio de urgencias del mismo hospital introduciría un sesgo debido a que este grupo de pacientes presentan un consumo de alcohol mayor que la población general de la que proceden los casos. Una manera de reducir este sesgo, es eligiendo los controles con diferentes enfermedades, de modo que los sesgos introducidos por enfermedades específicas tenderán a anularse entre sí.

Ejemplo 9.12. En el estudio caso-control italiano de cáncer de estómago mencionado en los Ejemplos 9.6 y 9.7, los controles se eligieron aleatoriamente de listas de población dentro de estratos de edad de 5 años y por sexo. Se tomó una muestra de un total de 1.423 controles de población, de los cuales se entrevistaron un total de 1.159 (81%) utilizando el mismo cuestionario estructurado que para los casos (Buiatti et al., 1989a,b). En la Tabla 9.3 se presenta el número de controles que se seleccionaron, los que se incluyeron en el estudio y las razones para la no participación por cada centro de reclutamiento.

Centro de reclutamiento	Seleccionados n (%)	Incluidos n (%)	Excluidos por	
			Rechazo n (%)	Mala salud n (%)
Cagliari	118 (100)	108 (91,5)	8 (6,8)	2 (1,7)
Cremona	61 (100)	51 (83,6)	5 (8,2)	5 (8,2)
Florence	547 (100)	440 (80,4)	74 (13,6)	33 (6,0)
Forli	291 (100)	259 (89,0)	20 (6,9)	12 (4,1)
Genoa	205 (100)	137 (66,8)	17 (8,3)	51 (24,9)
Imola	74 (100)	61 (82,4)	10 (13,5)	3 (4,1)
Siena	127 (100)	103 (81,1)	6 (4,7)	18 (14,2)
Total	1.423 (100)	1.159 (81,4)	140 (9,9)	124 (8,7)

^a Datos de Buiatti et al. (1989a).

Tabla 9.3.

Niveles de reclutamiento de controles y razones de no inclusión por centro de reclutamiento^a.

La elección de un grupo control adecuado es la parte más difícil del diseño de un estudio caso-control. Algunos estudios utilizan más de un tipo de grupo control. Las conclusiones de un estudio se potencian si se obtienen resultados similares con cada uno de los grupos control.

Una vez determinados el origen y el número de grupos control para un estudio, es necesario decidir cuántos controles deben elegirse por caso. Este aspecto se considerará con detalle en el Capítulo 15, pero cuando el número de casos y controles disponibles es grande y el coste de obtener información de los dos grupos es comparable, la relación óptima control-caso es 1:1. Cuando el número de casos disponibles para el estudio es bajo, o cuan-

do el coste de obtener información es mayor para los casos que para los controles, la relación control-caso puede modificarse para asegurar que el estudio pueda detectar un efecto, si realmente existe alguno (es decir, que el estudio tenga el poder estadístico necesario). Cuanto mayor sea el número de controles por caso, mayor será el poder del estudio (para un número de casos dado). No obstante, en general no está justificado aumentar esta relación por encima de 4:1 porque el aumento de poder estadístico con cada control adicional más allá de esta proporción tiene una magnitud limitada. Los aspectos del tamaño de la muestra (y el concepto de 'poder estadístico' de un estudio) se comentarán en el Capítulo 15.

Al igual que para todos los casos, es importante obtener información sobre las razones por las que los controles no participan y, siempre que sea posible, obtener información adicional sobre sus características sociodemográficas (p.e. sexo, edad, situación socioeconómica) (como en el Ejemplo 9.12).

9.3.3 Esquemas de muestreo para controles

Si la fuente de casos incidentes es una cohorte cerrada con un periodo de seguimiento fijo (como es el caso en estudios caso-control anidados), los controles pueden elegirse de tres formas diferentes tal y como se muestra en el Ejemplo 9.13.

En este ejemplo, la primera opción para los investigadores es seleccionar la muestra de los controles a partir de la población todavía a riesgo al final del estudio, es decir a partir de los sujetos que todavía no tenían la enfermedad al final del periodo de seguimiento (A en la Figura 9.1). En este diseño cada mujer puede ser un caso o un control, pero no los dos.

La segunda opción es seleccionar la muestra de los controles a partir de aquellos que todavía están a riesgo en el momento que se diagnostica cada caso, es decir, los controles están *apareados por tiempo* con los casos (véase Apartado 9.3.4) (B en la Figura 9.1). En este diseño de muestreo, un sujeto elegido inicialmente como control puede convertirse en caso en una fase posterior. Lo contrario no puede ocurrir, ya que una vez una mujer ha adquirido el cáncer de endometrio deja de estar a riesgo, y por tanto no será elegible para la selección como control. Los sujetos elegidos como controles que se convierten en casos deben mantenerse como controles y como casos, respectivamente, en los grupos adecuados.

En tercer lugar, los controles pueden ser una muestra de los que estaban a riesgo al inicio del estudio (C en la Figura 9.1). Este tipo de estudios se denomina '*estudios de caso-cohorte*'. Debido a que el grupo control refleja la población total y no sólo aquellos que no adquirieron la enfermedad, una mujer determinada como caso también puede elegirse como control y vice versa. Estas mujeres deben incluirse en el estudio por duplicado como casos y como controles.

Si la fuente de casos incidentes es una población dinámica, como en la mayoría de los estudios caso-control de hospitales y los basados en la población, suele ser difícil establecer la población exacta a riesgo al inicio y al final del periodo de seguimiento y el método preferido es elegir para cada

Ejemplo 9.13. Supongamos que se siguió una cohorte de 200.000 mujeres sanas durante diez años para evaluar la relación entre las variables de estilo de vida y el riesgo de desarrollar diferentes tipos de cáncer, y que a un total de 60 mujeres se les diagnosticó cáncer de endometrio durante el periodo de seguimiento. Supongamos también que los investigadores decidieron realizar un estudio caso-control anidado dentro de esta cohorte para evaluar si el uso de anticonceptivos orales protegía frente al cáncer de endometrio. Los investigadores podrían seleccionar la muestra de los controles de tres formas diferentes como se indica en la Figura 9.1: 1) a partir de los que todavía no tenían la enfermedad al final del periodo de seguimiento (situación A); 2) a partir de los que todavía estaban a riesgo en el momento del diagnóstico de cada caso (situación B); 3) a partir de aquellos que estaban a riesgo al inicio del estudio (situación C).

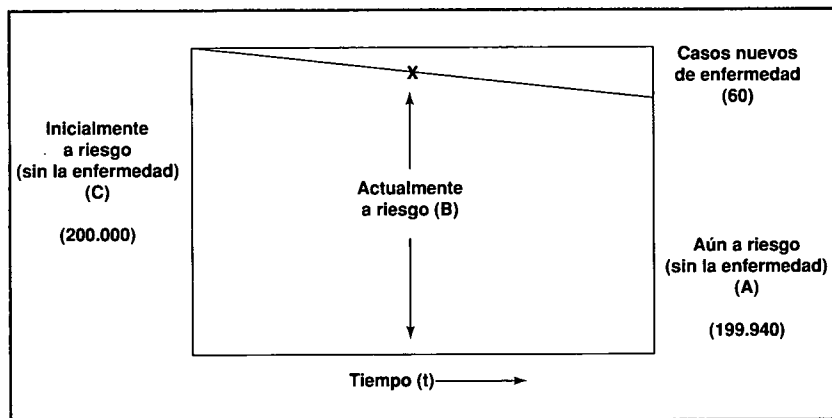


Figura 9.1.

Esquemas de muestreo para controles cuando la fuente de casos incidentes es una cohorte cerrada con un seguimiento establecido (x= caso incidente).

caso incidente uno o más controles entre los sujetos miembros de la misma población y que todavía están a riesgo de desarrollar la enfermedad en el momento del diagnóstico del caso.

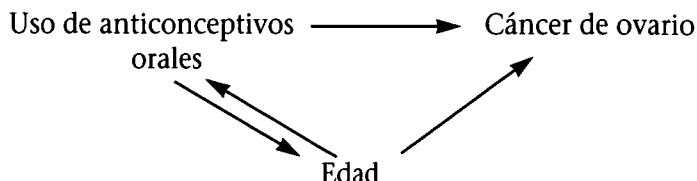
Como veremos más adelante en este capítulo (Apartado 9.5), la medida relativa específica del efecto (razón de tasas, razón de riesgos u *odds* (de la enfermedad) *ratio*) que pueden estimarse a partir de un estudio caso-control depende del tipo de diseño de muestreo utilizado en la selección de los controles.

9.3.4 Apareamiento

El *apareamiento individual* se refiere al procedimiento en el que uno o más controles se eligen para cada caso en base a la similitud respecto a determinadas características diferentes de la exposición en estudio. Debido a que los casos y controles son similares en las variables de apareamiento, su diferencia respecto al estado de la enfermedad puede atribuirse a diferencias en otros factores. No obstante, es importante que el *apareamiento se limite a factores de confusión y no se realice para la exposición a estudio*. Las ca-

racterísticas elegidas normalmente para aparear son las que se sabe son potentes *variables de confusión*. Las variables de apareamiento habituales son edad, sexo y raza, pero también pueden serlo el lugar de residencia o la situación socioeconómica.

Supongamos que estamos interesados en examinar la relación entre el uso actual de anticonceptivos orales y el cáncer de ovario.



En este ejemplo, es adecuado aparear por edad ya que la edad está asociada con la exposición de interés (uso actual de anticonceptivos orales) y es un factor de riesgo independiente de cáncer de ovario. En otras palabras, la edad es un factor de confusión. La falta de apareamiento, o de otro tipo de control, según la edad daría lugar a una evaluación sesgada del efecto del uso de anticonceptivos orales.

Cuando los controles se eligen de modo que sean similares a los casos para una característica y cuando esta similitud tiende a enmascarar la asociación de la enfermedad con la exposición de interés, se dice que los casos y controles están *apareados excesivamente*. Esto puede producirse cuando los controles están apareados con los casos según una característica que forma parte del proceso a través del cual la posible causa de interés da lugar a la enfermedad. Imaginemos que un estudio caso-control se realizó en el oeste de África para investigar el papel del virus de la hepatitis B en la etiología del cáncer de hígado, en el que los controles estaban apareados con los casos basándose en una historia de enfermedad hepática crónica.

Virus de la hepatitis B → Enfermedad hepática crónica → Cáncer de hígado

Si la enfermedad hepática crónica está en la vía causal entre la infección por el virus de la hepatitis B y el cáncer de hígado, el apareamiento según ésta daría lugar a una infraestimación del efecto del virus sobre la aparición del cáncer de hígado ya que los controles se habrían vuelto similares a los casos en relación a esta variable.

Otra forma de apareamiento excesivo se refiere a aparear según una variable que está correlacionada con la exposición de interés pero no es un factor de riesgo independiente para la enfermedad en estudio (y por tanto no puede ser un factor de confusión) y no está en la vía causal. Por ejemplo, supongamos que se quiere examinar la relación entre fumar y cáncer de pulmón en una población en la que los niveles de fumar están correlacionados de manera positiva con el consumo de alcohol, es decir, cuanto más bebe alguien, más probabilidad tiene de fumar. En este ejemplo, el

apareamiento según el consumo de alcohol daría lugar a un apareamiento excesivo debido a que los controles serían similares a los casos no sólo respecto al consumo de alcohol, sino también respecto a los hábitos de fumar, que es la exposición de interés en este estudio.

Por tanto, debe prestarse atención al determinar el número de variables elegidas para el apareamiento, incluso cuando no existan limitaciones prácticas. Si el papel de una variable es dudoso, la estrategia preferible no es aparear sino ajustar por ella en el análisis estadístico (Capítulos 13-14).

En la mayoría de los estudios caso-control, normalmente hay pocos casos y muchos controles potenciales para elegir (o seleccionar la muestra). En la práctica, cada caso se clasifica por las características que no son de interés directo y se realiza una búsqueda de uno o más controles con la misma serie de características.

Ejemplo 9.14. *El adenocarcinoma de vagina en mujeres jóvenes era un diagnóstico muy infrecuente hasta la notificación de varios casos tratados en el Hospital Vincent Memorial (en Boston MA. EE.UU.) entre 1966 y 1969. El diagnóstico inusual de este tumor en ocho pacientes jóvenes condujo a la realización de un estudio caso-control para identificar los posibles factores etiológicos. Para cada una de las ocho pacientes con carcinoma de vagina, se eligieron cuatro controles femeninos apareados por la fecha de nacimiento con una variación de cinco días y en el mismo tipo de servicio hospitalario (público o privado) que el caso según la información obtenida de los registros de nacimiento del hospital en el que había nacido cada paciente. Se entrevistaron todas las madres con un cuestionario estándar mediante entrevista personal efectuada por un entrevistador preparado específicamente para conducir estas entrevistas (Herbst et al., 1971).*

En el Ejemplo 9.14 los casos y controles se aparearon individualmente según el año de nacimiento (y por tanto, edad), hospital de nacimiento y tipo de servicio hospitalario (público frente a privado). Si los factores no son demasiado numerosos y existe una reserva importante de personas a partir de la cual pueden elegirse los controles, el apareamiento individual caso-control puede realizarse fácilmente. No obstante, si se consideran varias características o niveles y no existen muchos más controles potenciales que los casos, el apareamiento puede ser difícil y es probable que no se encuentren controles para algunos casos. Además, cuando se aparean los casos y los controles según alguna característica elegida, ya no puede estudiarse la influencia de esta característica sobre la enfermedad. En realidad el número de características para las cuales el apareamiento es deseable y práctico es bastante pequeño. Suele ser adecuado aparear casos y controles sólo según características tales como edad, sexo y raza, cuya relación con la enfermedad en estudio es conocida.

Como alternativa al apareamiento individual, puede *aparearse por frecuencia (o por grupo)*. Esto implica la selección de controles de modo que una

proporción similar de los casos se encuentre en las diferentes categorías definidas por la variable de apareamiento. Por ejemplo, si el 25% de los casos son hombres de 65-75 años, se tomará un 25% de los controles con características similares. El apareamiento por frecuencia dentro de categorías bastante amplias es suficiente en la mayoría de los estudios.

9.4 Determinación de las exposiciones

Los datos sobre las exposiciones relevantes pueden obtenerse mediante una entrevista personal, postal o telefónica, examinando los registros médicos, ocupacionales u otros, o tomando muestras biológicas. Para cualquier método elegido, es fundamental asegurar que la información obtenida no esté sesgada, es decir, no esté influenciada por el hecho de que un individuo sea un caso o un control. Lo ideal sería que el investigador o el entrevistador desconociera la hipótesis en estudio y la situación caso/control de los sujetos del estudio. En la realidad, esto puede ser difícil de cumplir, pero debe hacerse todo lo posible para asegurar una obtención no sesgada de los datos y reducir el *sesgo del observador*. En estudios multicéntricos será necesario un esfuerzo especial para asegurar la estandarización de las técnicas de obtención de datos en los diferentes centros participantes.

También puede producirse un sesgo cuando la validez de la información de exposición proporcionada por los sujetos difiere en los casos y controles (*sesgo del entrevistado*). Es probable que los sujetos con una enfermedad grave hayan estado pensando mucho sobre las posibles causas de su enfermedad de modo que los casos pueden tender a dar respuestas que se ajusten con lo que creen (o piensan que es aceptable decir) es la causa de su enfermedad. Este tipo particular de sesgo del entrevistado se denomina *sesgo de memoria*. El sesgo del entrevistado puede reducirse si los participantes no conocen las hipótesis en estudio y, si es posible, asegurando que los casos y controles tengan unos incentivos similares a recordar los acontecimientos del pasado. Estos aspectos se comentarán más adelante en el Capítulo 13.

9.5 Análisis

El análisis de datos de estudios caso-control depende de su diseño. Los estudios apareados individualmente requieren un tipo diferente de análisis que los estudios no apareados (o apareados por frecuencia).

9.5.1 Estudios no apareados (y apareados por frecuencia)

El primer paso en el análisis de un estudio caso-control no apareado es construir una tabla que muestre la frecuencia de las variables de interés por separado para casos y controles. La frecuencia de algunas de estas variables en los controles puede ayudar a evaluar la probabilidad de que representen a la población de la que proceden los casos. Por ejemplo, en el Ejemplo 9.15 la distribución de los años de escolarización, número de hijos, hábito tabáquico, etc. en el grupo control de este estudio basado en la población puede compararse con estadísticas generales de población o con resultados de otros estudios realizados en las mismas zonas.

Ejemplo 9.15. En el estudio caso-control de cáncer de cuello uterino descrito en el Ejemplo 9.2, se examinaron entre casos y controles la distribución de variables que se sabe son factores de riesgo para el cáncer de cuello uterino. La Tabla 9.4 muestra los resultados para algunas de estas variables.

	España		Colombia	
	Casos Nº (%)	Controles Nº (%)	Casos Nº (%)	Controles Nº (%)
Edad (años)				
<30	7 (2,8)	7 (2,9)	15 (8,1)	10 (6,7)
30-39	41 (16,4)	39 (16,4)	48 (25,8)	30 (20,1)
40-44	30 (12,0)	27 (11,3)	27 (14,5)	24 (16,1)
45-54	61 (24,4)	58 (24,4)	50 (26,9)	40 (26,9)
55+	111 (44,4)	107 (45,0)	46 (24,7)	45 (30,2)
Escolarización				
Alguna vez	162 (64,8)	179 (75,2)	155 (83,3)	140 (94,0)
Nunca	88 (35,2)	59 (24,8)	31 (16,7)	9 (6,0)
Número de hijos				
0-2	102 (40,8)	136 (57,1)	41 (22,0)	45 (30,2)
3-5	119 (47,6)	86 (36,1)	62 (33,3)	53 (35,6)
6+	29 (11,6)	16 (6,7)	83 (44,6)	51 (34,2)
Número de compañeros sexuales				
0-1	189 (75,6)	218 (91,6)	76 (40,9)	87 (58,4)
2-5	48 (19,2)	16 (6,7)	77 (41,4)	58 (38,9)
6+	13 (5,2)	4 (1,7)	33 (17,7)	4 (2,7)
Anticonceptivos orales				
Nunca	182 (72,8)	175 (73,5)	109 (58,6)	95 (63,8)
Alguna vez	64 (25,6)	53 (22,3)	77 (41,4)	53 (35,6)
No se sabe	4 (1,6)	10 (4,2)	0 (0,0)	1 (0,6)
Hábito tabáquico				
Nunca	184 (73,6)	198 (83,2)	92 (49,5)	90 (60,4)
Alguna vez	66 (26,4)	40 (16,8)	94 (50,5)	59 (39,6)

^a Datos de Bosch *et al.* (1992)

Tabla 9.4.

Distribución de algunos indicadores socio-económicos y reproductores de los casos de cáncer de cuello uterino y de los controles, en cada uno de los países participantes^a.

En el Ejemplo 9.15, las distribuciones de algunas de las variables conocidas como factores de riesgo para cáncer de cuello uterino se corresponden con las observadas en otros estudios en el sentido que los casos presentan un nivel cultural menor, más hijos y un mayor número de compañeros sexuales que los controles. También era más probable que hubieran utilizado alguna vez anticonceptivos orales o que fueran fumadoras.

En un estudio no apareado, el número de casos y controles expuestos y no expuestos al factor a estudio pueden representarse en una tabla 2×2 como se muestra en la Tabla 9.5:

	Expuestos	No expuestos	Total
Casos	<i>a</i>	<i>b</i>	n_1
Controles	<i>c</i>	<i>d</i>	n_0
Total	m_1	m_0	N

Tabla 9.5.

Representación de una tabla 2×2 que muestra los datos de un estudio caso-control no apareado.

En el Apartado 4.2.2, se presentaron las tres medidas de la incidencia (riesgo, *odds* de enfermedad y tasa) que pueden estimarse a partir de un estudio de cohorte. Estas tres medidas utilizan el mismo numerador –el número de casos nuevos que aparecieron durante el periodo de seguimiento –pero denominadores diferentes. El riesgo toma como denominador el número de personas a riesgo al inicio del seguimiento; la *odds* de enfermedad toma aquellos que no tenían la enfermedad al final del seguimiento; y la tasa utiliza el total de persona-tiempo a riesgo que tiene en cuenta el tiempo exacto en que se produjeron los casos. La comparación de estas medidas de incidencia en los expuestos respecto a los no expuestos da lugar a tres medidas diferentes del efecto relativo: la razón de riesgos, la *odds* (de enfermedad) *ratio* y la razón de tasas, respectivamente (Apartado 5.2.1).

En estudios caso-control, no puede estimarse directamente la incidencia de la enfermedad en los expuestos y los no expuestos ya que las personas se eligen basándose en si presentan o no la enfermedad de interés, y no basándose en su situación de exposición⁴. No obstante, puede calcularse la *odds* de la exposición en los casos y en los controles:

$$\text{Odds de exposición en los casos} = a/b$$

$$\text{Odds de exposición en los controles} = c/d$$

La *odds* (de exposición) *ratio* puede calcularse como

$$\text{Odds (de exposición) ratio} = \frac{\text{Odds de exposición en los casos}}{\text{Odds de exposición en los controles}} = \frac{a/b}{c/d}$$

Ejemplo 9.16. En el estudio caso-control presentado en el ejemplo anterior, se examinó el riesgo de cáncer de cuello uterino respecto a la educación (escolarización). En este análisis se reunieron los datos de España y Colombia (Tabla 9.6) (Bosch et al., 1992).

Tabla 9.6. Escolarización de casos de cáncer de cuello uterino y de controles^a.

	Escolarización		Total
	Nunca ('expuestos')	Alguna vez ('no expuestos')	
Casos de cáncer de cuello uterino	119 (a)	317 (b)	436 (n ₁)
Controles	68 (c)	319 (d)	387 (n ₀)
Total	187 (m₁)	636 (m₀)	823 (N)

^a Datos de Bosch et al. (1992)

Odds ratio = (119 / 317) / (68 / 319) = 1,76

Intervalo de confianza al 95% = 1,24–2,46

Test de $\chi^2 = 11,04$, 1 g.l.; Valor de $P = 0,0009$

(Los intervalos de confianza y las pruebas estadísticas de la *odds ratio* se calcularon como se indica en el Anexo A16.1)

⁴ Se pueden realizar cálculos indirectos en estudios caso-control basados en la población (véase A16.1).

Se puede demostrar algebraicamente que la *odds* (de exposición) *ratio* obtenida en un estudio caso-control proporciona una estimación no sesgada de una de las tres medidas del efecto que puede obtenerse de un estudio de cohorte, dependiendo del esquema de muestreo utilizado para elegir los controles (Apartado 9.3.3). Si los controles se eligen entre todos los que inicialmente están a riesgo, el estudio caso-control estimará directamente la *razón de riesgos*. Si los controles se eligen de aquellos que no presentan la enfermedad al final del seguimiento, el estudio estimará la *odds* (de enfermedad) *ratio*. Si los controles se eligen entre los que todavía están a riesgo en el momento en que se identifica cada caso, el estudio proporcionará una estimación no sesgada de la *razón de tasas*. En este último caso, el análisis debe considerar el hecho de que los casos y los controles están *apareados* respecto al tiempo. (Un análisis no apareado también proporcionará una estimación no sesgada de la razón de tasas si las tasas de adquisición de la enfermedad permanecen constantes en el tiempo entre las poblaciones expuestas y las no expuestas, y los números totales a riesgo siguen relativamente constantes en las dos poblaciones).

Como vimos en el Apartado 5.2.1, cuando la enfermedad es rara (p.e. cáncer) los casos constituyen una fracción negligible en relación a la población de referencia. El número de personas a riesgo en un estudio de cohorte permanece prácticamente constante en el tiempo y por tanto las tres medidas del efecto proporcionan resultados similares. Por consiguiente, los tres esquemas de muestreo utilizados para elegir a los controles en un estudio caso-control también proporcionarán resultados similares. No obstante, si la enfermedad es *frecuente*, con el mismo esquema de muestreo se obtendrán resultados diferentes y la elección del más adecuado dependerá del problema particular tratado en el estudio (Smith *et al.*, 1984; Rodrigues y Kirkwood, 1990).

En términos estrictos, la *odds ratio* obtenida en un estudio caso-control indica cuántas veces es mayor (o menor, si la exposición está asociada a un riesgo reducido) la probabilidad de que los casos hayan estado expuestos al factor en estudio en comparación con los controles. En el Ejemplo 9.16, los casos de cáncer de cuello uterino presentaron una probabilidad del 76% mayor que los controles de no haber asistido nunca al colegio. Debido a que la *odds ratio* obtenida de un estudio caso-control proporciona una estimación de una de las tres medidas del efecto que puede calcularse de un estudio de cohorte, también podemos interpretarlo como una indicación de la probabilidad de desarrollar la enfermedad en los individuos expuestos respecto a los no expuestos. En nuestro ejemplo, la *odds ratio* indica que las mujeres que nunca asistieron al colegio presentaban una probabilidad del 76% mayor de desarrollar cáncer de cuello uterino en comparación con las que asistieron.

Al igual que en otros tipos de estudios, las inferencias sobre la asociación entre una enfermedad y un factor se potencian considerablemente si se dispone de información para justificar un gradiente entre el nivel (o intensidad) de exposición y el riesgo de la enfermedad en cuestión. Pueden calcu-

larse las *odds ratios* para cada nivel de la exposición. El método general es tratar los datos como una serie de tablas 2 × 2, comparando a los casos y los controles según diferentes niveles de exposición y calcular el riesgo en cada nivel.

Tabla 9.7.
Número de parejas sexuales a lo largo de la vida entre casos y controles^a.

Ejemplo 9.17. En el estudio caso-control de cáncer de cuello uterino realizado en Colombia y España y descrito en los Ejemplos 9.2 y 9.15, se examinó el riesgo de desarrollar cáncer de cuello uterino respecto al número de parejas sexuales en la vida (Bosch et al., 1992). En este análisis se reunieron los datos de España y Colombia (Tabla 9.7).

Número de parejas sexuales	Casos de cáncer de cuello uterino	Controles	Odds ratio (intervalos de confianza al 95%)
0-1 ^b	265	305	1.0 ^c
2-5	125	74	1.94 (1.39-2.70)
6+	46	8	6.62 (3.07-14.27)

^a Datos de Bosch et al. (1992)

^b Considerado como la categoría basal (de referencia).

^c Test de χ^2 de tendencia lineal = 39,48; 1 g.l.; $P < 0,00001$

(Intervalos de confianza y prueba de χ^2 de tendencia lineal de los *odds ratios* calculados como se muestra en el Anexo 6.1.)

En el Ejemplo 9.17, existe una tendencia a aumentar el riesgo de cáncer de cuello uterino con el número de parejas sexuales. Las *odds ratios* para cada categoría de exposición se calcularon como sigue:

	Número de parejas sexuales	
	2-5 ('expuestos')	0-1 ('no expuestos')
Casos de cáncer de cuello uterino	125	265
Controles	74	305

$$Odds\ ratio = (125 / 265) / (74 / 305) = 1,94$$

$$Odds\ ratio = (46 / 265) / (8 / 305) = 6,62$$

	Número de parejas sexuales	
	6+ ('expuestos')	0-1 ('no expuestos')
Casos de cáncer de cuello uterino	46	265
Controles	8	305

Se pueden utilizar técnicas estadísticas especiales para ajustar los factores potenciales de confusión en el análisis. Éstos se comentarán en los Capítulos 13 y 14. Una de estas técnicas se utilizó para examinar la asociación entre escolarización y el riesgo de desarrollar cáncer de cuello uterino observada en el Ejemplo 9.16. La *odds ratio cruda* fue 1,8. Tras considerar las diferencias en la edad, centro participante, presencia o ausencia del virus del papiloma humano, número de parejas sexuales, educación, edad al te-

ner el primer hijo e historia de cribajes anteriores entre casos y controles, la *odds ratio ajustada* resultante fue de 2,5 (intervalo de confianza al 95%: 1,6-3,9). Por tanto, la asociación entre no haber asistido nunca a un colegio y el cáncer de cuello uterino observada en los análisis crudos no podía explicarse por diferencias en la distribución de ninguno de estos factores (de hecho, la *odds ratio* ajustada fue superior que la *odds ratio* cruda) (véanse más comentarios en el Apartado 13.2 y el Capítulo 14).

9.5.2 Estudios apareados por individuo

Los estudios apareados por individuo requieren un tipo de análisis específico, en el que la tabla 2x2 adquiere una forma diferente. Consideremos la situación más simple en la que sólo existe un control por caso. La situación de los casos respecto a la presencia o ausencia de la exposición de interés se tabula de forma cruzada frente a la situación de la exposición de sus controles correspondientes (Tabla 9.8).

		Controles		Total
		Expuestos	No expuestos	
Casos	Expuestos	<i>r</i>	<i>s</i>	<i>a</i>
	No expuestos	<i>t</i>	<i>u</i>	<i>b</i>
Total		<i>c</i>	<i>d</i>	<i>N/2</i>

Tabla 9.8.

Representación de una tabla 2x2 con datos de un estudio caso-control apareado por individuo (razón control-caso = 1:1).

En esta tabla, *r*, *s*, *t*, *u* representan el número de pares en que:

r = casos expuestos y controles expuestos (+ +)

s = casos expuestos pero controles no expuestos (+ -)

t = casos no expuestos y controles expuestos (- +)

u = casos no expuestos y controles no expuestos (- -)

Los totales marginales (*a*, *b*, *c*, *d*) de esta tabla corresponden a las entradas en las celdas de la tabla para los estudios no apareados. El total para toda la tabla es *N/2* pares, donde *N* representa el número total de individuos.

La *odds ratio* apareada puede calcularse como:

$$\text{Odds ratio} = s/t \text{ (Si } t \text{ no es igual a } 0)$$

Este cálculo de la *odds ratio* considera sólo los pares discordantes. Por intuición puede explicarse: los pares en los que caso y control estuvieron expuestos o en los que los dos estuvieron no expuestos, no proporcionan información sobre la relación entre la exposición y la enfermedad (Ejemplo 9.18).

El análisis es más complejo que el mostrado aquí si existe más de un control por caso (véase Breslow y Day (1980), Capítulo 5).

Ejemplo 9.18. En Canadá se realizó un estudio caso-control para evaluar si los edulcorantes artificiales, especialmente la sacarina, aumentaban el riesgo de cáncer de vejiga. Los casos recién diagnosticados de cáncer de vejiga que se produjeron entre residentes de las provincias de Columbia británica, Nueva Escocia y Terranova entre abril de 1974 y junio de 1976, se identificaron mediante registros de cáncer provinciales y la colaboración de patólogos y urólogos. Se evaluaron un total de 821 casos elegibles y 632 de éstos se entrevistaron personalmente en sus casas utilizando un cuestionario estructurado. Las razones por las cuales no pudieron obtenerse algunas entrevistas incluyeron muerte (56), negativa (65), demasiado enfermo para ser entrevistado (25) y negativa por el médico tratante a que la entrevista tuviera lugar (34). La mayoría de las entrevistas se realizaron dentro de los 3 meses del diagnóstico y todas dentro de los 6 meses. Para cada caso, se entrevistó un individuo apareado según sexo, edad (dentro de 5 años) y residencia en el vecindario (Howe et al., 1977). El resultado principal se presenta en la Tabla 9.9.

Tabla 9.9.
Frecuencia de exposición a edulcorantes artificiales en 632 casos de cáncer de vejiga y sus controles apareados^a.

		Controles		Total
		Expuestos	No expuestos	
Casos	Expuestos	468 (r)	87 (s)	555 (a)
	No expuestos	73 (t)	4 (u)	77 (b)
Total		541 (c)	91 (d)	632 (N/2)

^a Datos de Howe et al. (1977)

Odds ratio apareada = $87/73 = 1,19$

Intervalo de confianza al 95% para la odds ratio apareada = 0,86–1,65

Text de χ^2 de McNemar = 1,23; $P = 0,27$.

(El cálculo de los intervalos de confianza y las pruebas de significación para los estudios caso-control apareados se presentan en Breslow y Day (1980).)

9.6 Interpretación de los resultados

Los estudios caso-control son adecuados para estudiar enfermedades de larga latencia debido a que no requieren un seguimiento prolongado. También son adecuados para estudiar enfermedades raras ya que un estudio de cohorte prospectivo requeriría la inclusión de un número muy grande de individuos y un periodo de seguimiento prolongado para asegurar la inclusión de un número de casos suficiente.

No obstante, la interpretación de los estudios caso-control es menos sencilla que la de los estudios de cohorte y el investigador siempre debe considerar que el resultado podría haber aparecido como resultado de un sesgo de selección en la elección de los casos y controles, de un sesgo de información en la recopilación de los datos de exposición o debido a factores de confusión que no se tuvieron en cuenta correctamente.

El problema potencial más grave en los estudios caso-control es que los métodos utilizados para seleccionar los casos y controles pueden producir grupos que no sean realmente comparables.

Ejemplo 9.19. *La relación entre el uso de anticonceptivos orales por mujeres jóvenes y el riesgo de cáncer de mama se investigó en un estudio caso-control basado en la población realizado en el distrito de Los Angeles. Los casos eran pacientes con cáncer de mama confirmado con histología, diagnosticados por primera vez entre julio de 1972 y mayo de 1982, diagnosticados antes de los 37 años y sin una historia previa de enfermedad maligna. Se identificaron un total de 510 casos elegibles en el registro local del cáncer basado en la población, de los cuales 458 todavía estaban vivos en el momento del primer contacto. Los médicos responsables de cada caso autorizaron a ponerse en contacto con 393 (86%) de estas pacientes. De éstas, 26 no pudieron localizarse y 37 se negaron a ser entrevistados. Por tanto, se obtuvieron los cuestionarios completos de 330 pacientes. Dieciséis de estas pacientes se excluyeron más adelante debido a que no se encontró un control adecuado apareado individualmente (Pike et al., 1983).*

En el Ejemplo 9.19, un sesgo de selección podría haber afectado a los resultados de este estudio ya que sólo se incluyeron en el análisis final un 62% (314/510) de todas las pacientes elegibles. Las tasas de participación bajas pueden introducir un sesgo si los casos que utilizaron anticonceptivos orales hubieran sido más o menos propensos a participar en el estudio que los casos que no habían utilizado anticonceptivos orales. Si, por ejemplo, las mujeres que utilizaban anticonceptivos orales presentarían una forma menos agresiva de cáncer de mama que las que no los utilizaban y, por tanto, una mejor supervivencia, esto daría lugar a una sobreestimación del efecto de los anticonceptivos orales ya que una gran proporción de las muertes se habrían producido entre las que no los utilizaban.

La selección de un grupo control adecuado es uno de los problemas más difíciles en estudios caso-control. Los controles deben proceder de la misma población definida que los casos. El uso de controles basados en hospitales funciona sólo si los pacientes con enfermedades diferentes procedían de la misma población general (es decir, si los patrones de referencia son los mismos para la enfermedad en estudio y las enfermedades control) y si las enfermedades de los controles no están relacionadas con la exposición. En muchas situaciones es difícil estar seguro de que se cumplan estas condiciones. El uso de controles basados en la población evita estos problemas, pero todavía puede introducirse un sesgo de selección si los niveles de no participación son altos, bien porque algunos de los controles seleccionados no pueden localizarse o porque se niegan a participar (como en el Ejemplo 9.20). En este caso, puede que el grupo control no sea representativo de la población de la que proceden los casos.

Otro problema de los estudios caso-control es que normalmente es difícil obtener mediciones exactas de exposiciones en el pasado y el grado de exactitud y exhaustividad de estas mediciones puede ser diferente para casos y controles. Por ejemplo, en estudios caso-control puede aparecer un sesgo de memoria porque los pacientes con la enfermedad en estudio pue-

Ejemplo 9.20. *La posible asociación entre el uso de anticonceptivos orales y el riesgo de cáncer de mama en mujeres jóvenes (menores de 45 años) se investigó en un estudio caso-control poblacional realizado en Suecia y Noruega. En Noruega, donde es obligatoria la notificación de todos los diagnósticos de cáncer, se identificaron los casos a partir de registros de cáncer poblacionales. Se identificaron un total de 114 mujeres elegibles de las cuales 105 (92%) participaron. Para cada caso que aceptó participar, se eligieron 2 controles de un registro actualizado de la población nacional. Se envió una solicitud de participación a los controles potenciales. Si no se recibía una respuesta en cuatro semanas o si el control se negaba a participar, se elegía otro control. Nueve controles nunca se localizaron; 34 nunca respondieron la carta; 38 se negaron a participar; 4 estaban temporalmente ausentes y no se pudo entrar en contacto con ellos o presentaban trastornos mentales. Por tanto, para obtener 2 controles para cada caso fue necesario elegir 295 controles del registro de población. Sólo se entrevistaron un 72% de las mujeres inicialmente seleccionadas (Meirik et al., 1986).*

den tender a responder las preguntas más detenidamente que los sujetos control. Las comparaciones de las historias de la exposición obtenidas de los casos y controles con una fuente de información independiente (p.e. registros médicos) pueden ayudar a determinar la existencia o no de una diferencia sistemática en la memoria entre los casos y los controles.

Ejemplo 9.21. *En el estudio presentado en el Ejemplo 9.20, inicialmente se envió a casos y controles una carta introductoria y una descripción breve del objetivo y magnitud del estudio. Si aceptaban a participar se entrevistaban personalmente por mujeres entrevistadoras especialmente formadas (Meirik et al., 1986).*

En el Ejemplo 9.21, se explicó a las mujeres implicadas el objetivo de la investigación. Esto puede haber aumentado el sesgo de memoria, especialmente porque el estudio se efectuó en un momento de gran preocupación pública sobre los anticonceptivos orales y el cáncer de mama. Este problema podría haberse reducido en cierto grado no explicando la hipótesis del estudio a los sujetos.

La otra fuente potencial de sesgo en un estudio caso-control es el *sesgo de diagnóstico*. Por ejemplo, si las mujeres que utilizan anticonceptivos orales presentan una mayor tendencia que las que no los utilizan a realizar exámenes del pecho, a acudir a revisiones por un médico ó una enfermera, o a someterse a mamografías, puede introducirse un sesgo de diagnóstico. Por tanto, si se observa una relación positiva entre anticonceptivos orales y cáncer de mama en un estudio, ésta puede ser meramente debida al hecho de que las mujeres que toman anticonceptivos orales se controlan más y, por tanto, presentan una mayor probabilidad de que se les diagnostique

cáncer de mama que las que no los toman. Una manera de reducir el sesgo de diagnóstico es obtener información sobre la frecuencia de exámenes del pecho para cada uno de los sujetos en estudio, de modo que en el análisis pueden controlarse los efectos de una mayor frecuencia de revisiones de las mujeres que toman anticonceptivos orales.

Un estudio caso-control bien realizado que haya tenido en cuenta todos los aspectos metodológicos, puede proporcionar resultados válidos e informativos. Como se ha comentado anteriormente en este capítulo (Apartados 9.3.3 y 9.5), si los casos y controles se eligen independientemente de la exposición y los controles se eligen aleatoriamente a partir de una población de estudio bien definida de la que también proceden los casos, los resultados de un estudio caso-control proporcionan una estimación no sesgada de la medida del efecto que se obtendría en un estudio de cohorte equivalente. No obstante, es importante recordar que los estudios caso-control siempre presentan un potencial de sesgo y que cada estudio debe evaluarse individualmente para determinar si el sesgo afecta a los resultados. Con frecuencia, la dificultad reside en el hecho de que, aunque en un estudio caso-control determinado es fácil identificar las fuentes potenciales de error, raramente puede estimarse el impacto real que estos sesgos hayan podido tener sobre los resultados.

Una limitación importante de los estudios caso-control es que no pueden proporcionar estimaciones directas de la incidencia de la enfermedad en los expuestos y en los no expuestos (a no ser que sean basados en la población, véase Anexo 16.1). Por tanto, normalmente no puede calcularse el impacto absoluto de la exposición sobre la aparición de la enfermedad.

Los estudios caso-control no son adecuados para estudiar exposiciones raras debido a que muy pocos casos habrán estado expuestos, a no ser que una gran proporción de los casos totales de la enfermedad sea atribuible a esta exposición en particular (es decir, la fracción en exceso de la población es alta (véase Apartado 16.2.1)). Por ejemplo, la prevalencia de la exposición al amianto es rara en la población general y supone una pequeña proporción de los cánceres de pulmón. Por tanto, un estudio caso-control no sería adecuado para investigar la relación entre esta exposición y el cáncer de pulmón porque muy pocos casos habrían estado expuestos al amianto. No obstante, este diseño de estudio sería adecuado para investigar la relación entre el amianto y el cáncer de pleura debido a que esta exposición es la responsable de una gran proporción de los casos.

Por último, puede ser difícil de establecer la secuencia temporal entre la exposición y la enfermedad. Siempre debe considerarse la posibilidad de que la exposición sea el resultado (y no la causa) de la enfermedad (*causalidad inversa*). Por ejemplo, incluso si en un estudio caso-control se encuentra una asociación entre dieta y cáncer de estómago, existe una posibilidad de que la dieta sea una consecuencia en lugar de una causa del cáncer.

Lecturas adicionales

* En Lilienfeld y Lilienfeld (1979) se presenta una breve historia del desarrollo y uso de los estudios caso-control en epidemiología.

* Breslow y Day (1980) presentan una discusión muy completa del análisis de estudios caso-control en epidemiología del cáncer.

* En Miettinen (1976), Greenland y Thomas (1982), Smith *et al.* (1984) y Rodrigues y Kirkwood (1990) se presenta una discusión detallada de los esquemas de muestreo para la selección de controles.

Cuadro 9.1. Puntos esenciales

- Los estudios caso-control son estudios en los que se identifica un grupo de personas con la enfermedad de interés ('casos') y un grupo sin esta enfermedad ('controles') y se determina y compara la prevalencia (o nivel) de la exposición de interés en los dos grupos.
- Las principales ventajas de estos estudios son:
 1. Eficientes en tiempo y coste (por lo menos en comparación con estudios prospectivos de cohorte).
 2. Proporcionan la posibilidad de investigar una gran variedad de posibles factores de riesgo.
 3. Especialmente adecuados para investigar enfermedades raras o con un periodo de inducción prolongado.
- Los principales inconvenientes de estos estudios son:
 1. Difícil de encontrar un grupo control adecuado (sesgo de selección).
 2. Difícil de obtener medidas no sesgadas y precisas de exposiciones en el pasado (sesgo de exposición).
 3. La secuencia temporal entre exposición y enfermedad puede ser difícil de establecer (causalidad inversa).
 4. No son adecuados para investigar exposiciones raras (a no ser que la exposición sea responsable de una gran proporción de casos, es decir, la fracción etiológica de la población sea alta).
 5. No es posible obtener estimaciones de la incidencia de la enfermedad en los expuestos y los no expuestos a un posible factor de riesgo (excepto si el estudio se basa en la población).

Capítulo 10:

Estudios transversales

Los estudios transversales están destinados a determinar la prevalencia (o el nivel) de un atributo en particular, como una exposición o enfermedad específicas o cualquier otro acontecimiento relacionado con la salud, en una población definida en un *momento determinado en el tiempo*. Por ejemplo, se puede realizar un estudio transversal para estimar la prevalencia de la infección por hepatitis B, o la prevalencia de fumar, o la proporción de mujeres en edad fértil en periodo de lactancia en una población determinada en el momento de la encuesta (Ejemplo 10.1).

Ejemplo 10.1. *Las Encuestas sobre Fertilidad en el Mundo [World Fertility Surveys (WFS)] fueron estudios nacionales sobre el comportamiento reproductivo en humanos realizados en aproximadamente 40 países en vías de desarrollo y 20 países desarrollados a finales de los años 70. Entre otros aspectos del comportamiento reproductivo, estos estudios recopilaron información sobre la lactancia (Naciones Unidas, 1987). La Tabla 10.1 muestra el porcentaje de mujeres de 15-49 años en países seleccionados que estaban en periodo de lactancia en el tiempo en que se realizaron estos estudios.*

Región y país	Período de estudio	Tamaño de la muestra	Porcentaje de mujeres de 15-49 años actualmente en lactancia
<i>África</i>			
Egipto	1980	8.788	34,3
Ghana	1979-80	6.125	37,7
Kenya	1977-78	8.100	43,2
<i>América latina y Caribe</i>			
Colombia	1976	5.378	17,1
México	1976	7.310	19,8
Venezuela	1977	4.361	15,3
<i>Asia y Oceanía</i>			
Bangladesh	1975-76	6.513	51,1
Indonesia	1976	9.155	15,9
Pakistán	1975	4.996	40,5

^a Datos de Naciones Unidas (1987)

Tabla 10.1.

Proporción de mujeres de 15-49 años en países seleccionados que estaban en periodo de lactancia en el momento en que se realizaron las Encuestas de Fertilidad en el Mundo, 1975-1980^a.

Por tanto en este tipo de estudio, se identifica a los sujetos en un momento determinado del tiempo y se obtiene la información pertinente. Basándose en esta información, se clasifican según tengan o no el atributo de interés.

En algunos casos, los estudios transversales intentan ir más allá de proporcionar únicamente información sobre la frecuencia (o nivel) del atributo de interés en la población de estudio obteniendo información de los factores de riesgo potenciales. Por ejemplo, en un estudio transversal realizado para estimar la prevalencia de la hepatitis B en una población determinada también se pueden obtener datos sobre factores de riesgo potenciales para esta enfermedad tales como situación socioeconómica, uso de drogas por vía intravenosa, comportamiento sexual, etc.

Ejemplo 10.2. En las Encuestas de Fertilidad en el Mundo, se examinaron las prácticas de lactancia respecto a los factores socioeconómicos tales como la educación de la madre (Tabla 10.2).

Tabla 10.2.
Duración media de la lactancia (meses) según años de escolarización de la madre en países seleccionados. Encuestas de Fertilidad en el Mundo, 1975-1980^a.

Región y País (tamaño de la muestra)	Período de estudio	Años de escolarización de la madre			
		Cero	1-3	4-6	7+
<i>África</i>					
Egipto (8.788)	1980	21,2	19,5	16,3	10,2
Ghana (6.125)	1979-80	21,3	n/a.	19,2	15,7
Kenya (8.100)	1977-78	19,6	17,4	15,2	12,5
<i>América latina y Caribe</i>					
Colombia (5.378)	1976	11,9	11,4	8,3	5,3
México (7.310)	1976	12,9	10,9	8,3	3,8
Venezuela (4.361)	1977	11,6	10,0	6,7	3,5
<i>Asia y Oceanía</i>					
Bangladesh (6.513)	1975-76	34,4	30,4	n/a.	n/a.
Indonesia (9.155)	1976	28,4	27,0	24,7	13,7
Pakistán (4.996)	1975	22,0	n/a.	19,8	n/a.
^a Datos de Naciones Unidas (1987)					
n/a. = datos no disponibles debido al pequeño tamaño de las muestras.					

En el Ejemplo 10.2, se examinó la duración de la lactancia según años de escolarización de la madre. En todos los países en los que pudo realizarse la comparación, la duración de la lactancia disminuyó considerablemente al aumentar el nivel educacional de la madre.

Los estudios transversales también son útiles para evaluar prácticas, actitudes, conocimiento y creencias de una población respecto a un acontecimiento particular relacionado con la salud (Ejemplo 10.3). Los resultados de estos estudios no sólo indican la magnitud del problema en una pobla-

Ejemplo 10.3. En los EE.UU., en 1966, se realizó un estudio nacional para evaluar la prevalencia de tabaquismo y las actitudes y creencias sobre el uso de tabaco y otras variables relacionadas. El cuestionario incluía, entre otras, preguntas sobre los siguientes aspectos: hábito tabáquico (pasado y presente); intentos de dejar y/o reducir el hábito de fumar; estimaciones propias del comportamiento futuro sobre el hábito de fumar; creencias sobre la capacidad de cambiar y voluntad de cambiar; justificación del comportamiento del hábito de fumar; actitudes y creencias sobre fumar como un riesgo sanitario en general y para los sujetos en particular; gratificaciones derivadas del fumar y presiones sociales para la continuación o interrupción (Departamento de Sanidad, Educación y Bienestar de los EE.UU., 1969).

ción particular en un momento determinado, sino que también ayudan a diseñar medidas adecuadas de salud pública (p.e. campañas de educación sanitaria).

Los estudios transversales también son un método válido para obtener información sobre los patrones de morbilidad de la población, así como sobre la utilización de servicios sanitarios preventivos y curativos (Ejemplo 10.4). Sus resultados ayudan a los planificadores sanitarios a establecer prioridades sanitarias.

Ejemplo 10.4. El Programa Danfa de Salud Rural y Planificación Familiar [Danfa Comprehensive Rural Health and Family Planning Project], se inició para evaluar los sistemas sanitarios y de planificación familiar en el sur de Ghana. Como parte de este proyecto, se inició un estudio transversal familiar sobre morbilidad en la zona de estudio para proporcionar información de los patrones de enfermedad e incapacidad, cantidad de trabajo perdido y del uso de los servicios sanitarios durante el periodo de dos semanas anterior al estudio (Belcher et al., 1976).

10.1 Población diana y muestra de estudio

Al igual que para otros tipos de diseños epidemiológicos, los objetivos del estudio deben definirse claramente antes de su inicio. Esto debe incluir una definición clara del atributo de interés (enfermedad, exposición o cualquier otro acontecimiento relacionado con la salud) y de los factores de riesgo potenciales, así como una consideración clara de la *población diana*, es decir, la población a la cual se extrapolarán los principales resultados del estudio. Por ejemplo, supongamos que queremos efectuar un estudio sobre la dieta de los Adventistas del Séptimo Día (un grupo religioso que no come carne ni bebe alcohol). Será necesario decidir si deben incluirse niños, recién convertidos a la iglesia o los que la acaban de abandonar.

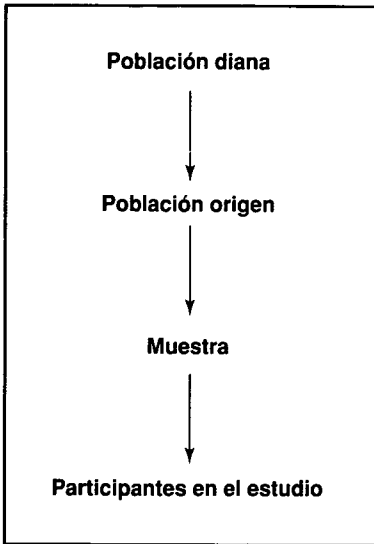


Figura 10.1. Diagrama que muestra la relación entre la población diana y los participantes en el estudio.

A continuación, es necesario identificar una *población origen* adecuada (Figura 10.1). Por razones prácticas y logísticas, la población origen generalmente está más limitada que la población diana. Por ejemplo, aunque nuestra población diana es todos los Adventistas del Séptimo Día, obviamente será imposible incluirlos a todos. La elección de la población origen debe determinarse por la definición de la población diana de estudio y por las limitaciones logísticas. Por razones logísticas podemos decidir realizar el estudio en California (EE.UU.) donde viven muchos Adventistas del Séptimo Día. Si esta población origen es lo suficientemente pequeña como para ser estudiada utilizando los recursos humanos y económicos disponibles, puede incluirse toda la población. Si la población de origen todavía es demasiado grande, debe seleccionarse una *muestra representativa*.

10.1.1 ¿Cómo seleccionar una muestra?

Con objeto de elegir una muestra de la población origen, es necesario decidir el *diseño de la muestra*, es decir, el método a utilizar para la selección de las unidades de muestreo de la población. Las muestras a veces se eligen por *criterio* (es decir, lo que el investigador considera una muestra 'equilibrada' o 'representativa') o por *conveniencia* (es decir, los sujetos obtenidos más fácilmente tales como voluntarios o personas que acuden a una clínica).

Ejemplo 10.5. En Seúl, Corea del Sur, se realizó un estudio para determinar la prevalencia de la infección por *Helicobacter pylori* en su población. La muestra estaba formada por adultos sanos asintomáticos y niños que visitaron un centro sanitario de cribaje en el Hospital Guro para una revisión rutinaria. La mayoría de los individuos proporcionados por este centro eran de clase media con escasa representación de pacientes privados y de las familias de clases socioeconómicas más bajas (Malaty et al., 1996).

Ninguno de estos métodos, proporciona garantías frente a la posibilidad de que pueda introducirse un sesgo de selección (consciente o inconsciente). Puede que algunas personas tengan una probabilidad aumentada de formar parte de la muestra, a consecuencia de lo cual, esta muestra no será representativa. Por ejemplo, los pacientes que acuden a una clínica pueden ser diferentes de los que no acuden (como en el Ejemplo 10.5).

El mejor método es el *muestreo aleatorio* en el que el azar es el único factor que determina quién se incluirá en la muestra, eliminando toda posibilidad de sesgo de selección.

Con objeto de obtener una muestra aleatoria de la población origen, es necesario tener un *marco de muestreo*, es decir, una relación completa de las unidades de muestreo de la población de estudio. La unidad de muestreo puede ser una persona individual, una familia o una escuela. Los registros electorales pueden ser un marco de muestreo adecuado para adultos pero no para niños. Si el marco de muestreo se basa en estadísticas oficiales, algunos grupos, tales como los inmigrantes recientes, las personas sin hogar y

los habitantes de barrios pobres, pueden estar infrarrepresentados. Para otros grupos como gitanos y otros grupos nómadas puede que no exista un marco de muestreo. En algunos países puede ser imposible enumerar a cada sujeto en la población de estudio. Como se verá más adelante en este capítulo (Apartado 10.1.4), en estas circunstancias pueden utilizarse técnicas especiales para asegurar la obtención de una muestra representativa.

10.1.2 Muestreo aleatorio simple

El tipo más elemental de muestra aleatoria es una muestra aleatoria simple, en la que cada unidad de muestreo tiene la misma posibilidad de ser seleccionada directamente de la población origen.

El primer paso es definir quienes son las unidades de la muestra, es decir, las personas o unidades (p.e. familias) a muestrear. Estas unidades se tienen que definir claramente en función de sus características particulares. El siguiente paso es tomar un marco de muestreo, es decir, una lista de todas las unidades de la muestra en la población origen. El marco de muestreo debe ser exhaustivo, completo y actualizado para mantener en un mínimo el sesgo de selección. Ejemplos generales de marcos de muestreo son las listas del censo, registros electorales, etc. Una vez identificado un marco de muestreo adecuado, debe asignarse un número a cada unidad. Si la población de origen está formada por 2.000 individuos, a cada uno se le debe asignar un único número entre 1 y 2.000. A continuación pueden utilizarse las tablas de números aleatorios para seleccionar una muestra aleatoria de un número determinado de unidades de muestreo a partir de toda la población de origen. En primer lugar, deben elegirse al azar un lugar de inicio en la tabla y una dirección. A continuación deben registrarse todos los dígitos secuenciales encontrados en la tabla, parando sólo cuando se alcance el tamaño de la muestra necesario (véase en el Apartado 7.9.2 una ilustración de cómo utilizar las tablas de números aleatorios). Como alternativa, pueden generarse secuencias de números aleatorios con una calculadora o un ordenador. La muestra estará formada por las unidades de muestreo que corresponden a estos números aleatorios.

La ventaja principal del muestreo aleatorio simple es que es relativamente sencillo en comparación con métodos más complejos (tales como los descritos en los Apartados 10.1.4 y 10.1.5). Su principal limitación es que sólo puede aplicarse cuando la población es relativamente pequeña y concentrada en una zona geográfica reducida y donde el marco de muestreo es completo.

En el Ejemplo 10.6, se eligió una muestra de mujeres por muestreo aleatorio simple en cada uno de los dos municipios. Este método fue adecuado debido a que allí había un marco de muestreo correcto (es decir, la lista informatizada) para cada zona y las dos poblaciones estaban concentradas en una zona geográfica relativamente pequeña.

Aunque normalmente no es factible utilizar el muestreo aleatorio simple para la selección de toda la muestra en un estudio grande, con frecuencia se utiliza para la selección final de las unidades del estudio (p.e. selección de familias en comunidades, una vez elegidas las comunidades) en esquemas más complejos como se describe en el Apartado 10.1.4.

Ejemplo 10.6. Se realizó un estudio transversal en muestras aleatorias de mujeres en una zona de alto riesgo de cáncer de cuello uterino (Nuuk, Groenlandia) y en una zona de bajo riesgo (Nykøbing Falster) de Dinamarca, para evaluar la prevalencia del virus del papiloma humano (VPH) y del virus del herpes simple (VHS). El Registro Central de Población Danés es un registro informatizado de todos los individuos que estaban vivos en 1968, nacidos o emigrantes a Dinamarca posteriormente e incluye información sobre el estado vital y la emigración. De este registro de la población se tomó una muestra al azar de 800 mujeres de 20-39 años, nacidas en Groenlandia y residentes en el municipio de Nuuk/Godthåb. Asimismo, del mismo registro se tomó una muestra al azar de 800 mujeres de 20-39 años, nacidas en Dinamarca y residentes en el municipio de Nykøbing Falster (Kjaer et al., 1988).

10.1.3 Muestreo sistemático

A veces puede ser más adecuado tomar una muestra sistemática que una muestra aleatoria simple. Para hacer esto, las unidades deben ordenarse en algún tipo de secuencia como en un directorio, una serie de tarjetas indexadas, casas en la calle o pacientes cuando llegan a una clínica. A continuación hay que decidir qué fracción de la población debe estudiarse. Supongamos, por ejemplo, que queremos seleccionar un tamaño de muestra de 40 sujetos entre una población de 200. Esto será una muestra de 1 a 5. Se elige un número al azar entre 1 y 5 –supongamos que fuera 2. Empezando desde el principio de la lista podemos seleccionar la unidad de muestreo número 2 y a continuación cada quinta unidad subsiguiente. La muestra incluirá unidades de 2, 7, 12, 17,... y así sucesivamente.

La principal ventaja del muestreo sistemático es la conveniencia. Normalmente proporciona una buena aproximación al muestreo aleatorio simple siempre que los intervalos no correspondan con cualquier patrón recurrente en la población de origen. Si aparecen características especiales en el marco de muestreo a intervalos regulares, se introducirá un sesgo. Consideremos que pasaría si la población estuviera formada por una serie de parejas casadas con los maridos listados siempre los primeros. Si se tomara cada cuarta persona daría lugar a una muestra formada exclusivamente por hombres si se empezara con el primer o tercer sujeto o formada exclusivamente por mujeres si se empezara con el segundo o el cuarto. Asimismo, cada casa número 20 en una lista de direcciones podría ser una casa en una esquina con características diferentes a las otras casas.

10.1.4 Muestreo en múltiples fases

En muchas situaciones no es factible o práctico tomar una muestra simple aleatoria o una muestra sistemática de toda la población de origen. Esto es debido a que no se dispone de un marco de muestreo y sería demasiado difícil crear uno, o porque la población está dispersada en una zona geográfica muy grande. Por ejemplo, no sería realista intentar tomar una muestra simple aleatoria de 200 personas de la población de todo un país.

Incluso aunque existiera un marco de muestreo adecuado, la mayoría de las muestras vivirían en comunidades diferentes alejadas entre sí, y los gastos implicados en ponerse en contacto con ellos serían prohibitivos. Una solución es utilizar el *muestreo en dos fases* como sigue:

1. En primer lugar se divide la población en *grupos*, Por ejemplo regiones, pueblos o distritos, y se hace una lista de estas *unidades de primera fase* (o unidades de muestra primaria).
2. A continuación se elige de esta lista de unidades de primera fase una muestra aleatoria.
3. En cada una de las unidades de primera fase elegidas, se toma un marco de muestreo de *unidades de segunda fase* (p.e. familias o individuos) y a continuación se elige una muestra aleatoria.

Ejemplo 10.7. *En China en 1983 se realizó un gran estudio transversal al final de la temporada de cosecha para proporcionar información sobre la dieta y estilo de vida. Para elegir a los participantes se utilizó un método de muestreo en múltiples fases.*

(1) *Se eligieron 65 distritos rurales (casi todos con poblaciones superiores a 100.000 en 1973-5) de un total de 2.392 distritos. La selección de los distritos no fue aleatoria sino que estaba diseñada para producir áreas geográficas con una amplia representación de tasas de cáncer de siete de los cánceres más frecuentes y una amplia dispersión geográfica.*

(2) *Se eligieron dos comunidades en cada uno de los 65 distritos, es decir, un total de 130 comunidades. Aunque la selección de las comunidades fue aleatoria, se decidió incluir únicamente a las comunidades que estuvieran a una distancia máxima de cuatro horas de viaje desde la comunidad de estudio al laboratorio del distrito, ocasionando la sustitución de seis comunidades.*

(3) *Se eligió al azar una brigada de producción en cada comunidad, es decir, un total de 130 brigadas de producción.*

(4) *En cada una de las 130 brigadas de producción se eligieron aleatoriamente dos equipos de producción, es decir, un total de 260 equipos de producción.*

(5) *Dentro de cada uno de los 260 equipos de producción, se eligieron al azar 25 familias de un registro oficial de residencias (proporcionando un total de 100 familias por distrito y 50 familias por comunidad).*

(6) *En cada familia se solicitó a un hombre o una mujer de 35-64 años que donara sangre y que completara un cuestionario sobre su dieta, su consumo de alcohol y tabaco y su historia reproductiva. Un total de 6.500 sujetos participaron en el estudio (Chen et al., 1990).*

Esta estrategia puede ampliarse a varias fases (*muestreo en múltiples fases*), como en el Ejemplo 10.7.

Si no existe el marco familiar y no puede crearse uno, debe utilizarse algún método de selección que asegure que la muestra sea lo más representativa posible. Esto normalmente implicará dos fases: un método de selección de una familia como punto de inicio y un procedimiento de selección de las siguientes familias. Una posibilidad es elegir un punto central en el pueblo, como el mercado o la plaza central; elegir una dirección al azar a partir de este punto (p.e. lanzando un lápiz al aire y viendo en que lado cae); contar el número de familias entre el punto central y el final del pueblo en esta dirección; elegir una de estas casas al azar como punto de inicio del estudio. Las familias restantes en la muestra deben elegirse de modo que proporcionen una cobertura amplia del pueblo. El método preciso utilizado no es demasiado importante, siempre que no de lugar a que todas las familias elegidas estén muy próximas entre sí y siempre que la norma para elegir las familias después de la primera sea simple e inequívoca, para eliminar la posibilidad de que los entrevistadores introduzcan un sesgo evitando algunas zonas.

Si el objetivo del estudio es obtener una estimación global de la prevalencia o del nivel de un atributo para toda la población diana, es lógico elegir las diferentes zonas geográficas de modo que la probabilidad de selección sea proporcional al tamaño de su población. Por ejemplo, una localidad con una población de 200.000 debe presentar una probabilidad 10 veces mayor de ser elegida respecto a una localidad de 20.000. Subsecuentemente debe tomarse un número similar de individuos o familias de los pueblos grandes y pequeños. Esta aproximación de *probabilidad proporcional al tamaño del muestreo* asegura que los individuos (o familias) en pueblos grandes y pequeños presentan la misma probabilidad de ser elegidos al principio del procedimiento de muestreo.

Las ventajas del muestreo en fases múltiples son obvias en términos de costes y tiempo. Por tanto, ¿deberíamos realizar todos los muestreos mediante selección de grupos de conveniencia? Un inconveniente de este método es que en muchas ocasiones puede que los grupos estén formados por individuos más homogéneos que la población general. Por ejemplo, es más probable que las personas que viven en el mismo vecindario o pueblo sean similares en cuanto a las características del estilo de vida. Si es así, los individuos en una muestra de vecinos proporcionan menos información que una muestra de tamaño similar obtenida de toda la población del estudio.

10.1.5 Muestreo aleatorio estratificado

Una *muestra aleatoria estratificada* implica la división de la población en subgrupos diferenciados según algunas características importantes, tales como sexo, edad o situación socioeconómica y la selección aleatoria de una muestra de cada subgrupo. Cada subgrupo se conoce como un *estrato* y se elige una muestra aleatoria diferente (simple o en fases múltiples) en cada uno.

En el Ejemplo 10.8, se formaron ocho estratos por edad y sexo (1, hom-

Ejemplo 10.8. *Las seroprevalencias de inmunoglobulinas G (IgG), M (IgM) y A (IgA), anticuerpos a Helicobacter pylori, se evaluaron mediante el ensayo de inmunoabsorción enzimática (ELISA) en un estudio realizado en la parte oeste de Copenhague. En 1982 se tomó del Registro de Población Danés una muestra aleatoria estratificada por edad y sexo formada por 4.807 hombres y mujeres nacidos en los años 1922, 1932, 1942 y 1952 (es decir, personas de 30, 40, 50 y 60 años) y residentes en la parte oeste de Copenhague. En el registro están incluidos todos los habitantes de Dinamarca (Andersen et al., 1996).*

bres nacidos en 1922; 2, hombres nacidos en 1932; 3, hombres nacidos en 1942; 4, hombres nacidos en 1952; 5, mujeres nacidas en 1922; 6, mujeres nacidas en 1932; 7, mujeres nacidas en 1942; 8, mujeres nacidas en 1952) y se tomó una muestra aleatoria dentro de cada uno.

Existen muchas situaciones en las que este tipo de muestreo es el más adecuado. A veces puede desearse disponer de resultados independientes para estratos diferentes y asegurar un tamaño de la muestra adecuado en cada uno. A su vez, podemos sospechar que la prevalencia variará entre estratos (como en el Ejemplo 10.8). Debido a que pueden utilizarse diferentes esquemas de muestreo en estratos diferentes, el muestreo aleatorio estratificado es especialmente adecuado cuando, por ejemplo, sólo se dispone de los marcos de muestreo para algunos subgrupos de la población (como en Ejemplo 10.9).

Ejemplo 10.9. *Los estudios de fertilidad mundial mencionados en los Ejemplos 10.1 y 10.2 utilizaron un diseño de muestreo general para elegir las muestras en los diferentes países participantes. En primer lugar, se identificó en cada país un marco de muestreo de unidades de área con unos límites bien definidos. En segundo lugar, se eligió aleatoriamente una muestra de estas unidades de área con una probabilidad proporcional al tamaño de sus poblaciones. En tercer lugar, se realizó una lista de viviendas o familias en las unidades de área elegidas. En cuarto lugar, se eligió al azar un número similar de viviendas o familias de cada unidad de área elegida y en estas familias se entrevistaron todas las mujeres que cumplían los criterios de inclusión. La estratificación también se utilizó en algunos países para asegurar que las zonas urbana y rural estaban correctamente representadas o debido a la necesidad de utilizar diseños de muestreo diferentes en algunas zonas geográficas. Por ejemplo, se disponía de listas de población para algunas áreas urbanas de las cuales se podían seleccionar las mujeres aleatoriamente sin la necesidad de realizar una lista de viviendas o familias (Scott y Harpham, 1987).*

10.1.6 Participantes en el estudio

No todos los sujetos elegidos en una muestra aleatoria acabarán participando en el estudio. Algunos sujetos se negarán a participar a pesar de todos los esfuerzos realizados, otros habrán muerto o se habrán trasladado de

la zona. Por tanto, los participantes normalmente son una parte de la muestra aleatoria inicial.

Ejemplo 10.10. *En el estudio transversal descrito en el Ejemplo 10.6, se eligió una muestra aleatoria de mujeres en cada una de las dos zonas geográficas de Dinamarca para evaluar (y comparar) la prevalencia del virus del papiloma humano (VPH) y del virus del herpes simple (VHS). Los investigadores describieron el método de reclutamiento de los participantes en el estudio en cada área como sigue:*

Se tomaron al azar 800 mujeres de 20-39 años, nacidas en Groenlandia y residentes en el municipio de Nuuk/Godthåb, a partir del Registro Central de Población informatizado del Reino Danés. De estas mujeres, 104 se habían trasladado del municipio y una había muerto antes de poder ponerse en contacto con ellas, quedando 695 para el estudio. [...] En los medios informativos locales (periódicos, radio) se proporcionó información general sobre el estudio tras lo cual se invitó a cada mujer [elegida al azar] mediante una carta personal, a visitar la clínica local. Dos o tres semanas después de la invitación inicial se enviaron recordatorios y por último se entró en contacto con las que no respondieron mediante un mensajero personal desde la clínica. De las 695 mujeres elegibles, 586 (84,3%) se incluyeron en el estudio, 93 (13,4%) no se localizaron y 16 (2,3%) se negaron a participar. La proporción relativamente alta de mujeres que no pudieron localizarse puede atribuirse a errores en el registro municipal de población y a deficiencias del servicio postal.

[...] También a partir del Registro Central de Población se tomó de la población femenina de 20-39 años una muestra aleatoria de 800 mujeres, nacidas en Dinamarca y residentes en el municipio de Nyøbing Falster. Catorce de estas mujeres se habían trasladado del municipio y una había muerto antes de la inclusión, quedando 785 mujeres para el estudio. En los medios informativos locales y nacionales se proporcionó información general sobre el estudio. A continuación se invitó a cada mujer [elegida al azar] mediante una carta personal a participar en el estudio y se programaron para una visita en el hospital local. Se enviaron recordatorios 2-3 semanas después de la invitación inicial y a las que no respondieron se les llamó por teléfono. Se incluyeron un total de 661 mujeres (84,2%), 58 (7,4%) no pudieron localizarse y 66 (8,4%) no quisieron participar (Kjaer et al., 1988).

Debe hacerse todo lo posible para asegurar un alto nivel de respuesta y de participación para reducir el sesgo de selección. Siempre debe indicarse el número de personas o familias entrevistadas, no sólo el número en la muestra original, de modo que puedan computarse los que no responden. ¿Qué nivel de respuesta debe considerarse aceptable en un estudio? Para una situación poco frecuente, una fracción de respuestas del 85% puede ser inaceptable, debido a que los pocos casos existentes en el 15% no examinado puede alterar los resultados considerablemente; por otro lado, en un estudio de un atributo relativamente frecuente este nivel de respuesta puede considerarse bueno.

Es probable que los participantes en cualquier estudio difieran en algunas de sus características con los que no responden. El aspecto importante es si esta diferencia introducirá un sesgo en el estudio y, si existe este sesgo, ¿en qué magnitud puede afectar a los resultados? Con objeto de evaluar el sesgo introducido por la falta de respuesta, es esencial intentar obtener alguna información sobre los individuos que inicialmente se negaron a participar o no pudieron localizarse. Son posibles dos aproximaciones. En primer lugar, puede tomarse una muestra pequeña al azar de los que no responden y hacer todo lo posible, incluyendo visitas a domicilio, para incitar su participación. Los resultados de esta muestra aleatoria indicarán la magnitud del sesgo global entre los que no responden. En segundo lugar, puede disponerse de información de todas las personas relacionadas en la población del estudio; a partir de esto será posible comparar los que responden y los que no responden respecto a características básicas tales como edad, sexo, residencia y situación socioeconómica.

Ejemplo 10.11. Consideremos otra vez el estudio descrito en el Ejemplo 10.10. La Tabla 10.3 presenta las distribuciones de edad de la población femenina total en las dos zonas geográficas elegidas, de las dos muestras aleatorias elegidas a partir de las mismas, de las mujeres que eran elegibles y de las que participaron realmente en el estudio.

Edad	Número de mujeres			
	Población femenina total	Muestra aleatoria	Mujeres elegibles	Participantes
<i>Nykøbing Falster (Dinamarca)</i>				
20–24	975 (27,9)	228 (28,5)	222 (28,3)	166 (25,1)
25–29	774 (22,2)	158 (19,8)	153 (19,5)	132 (20,0)
30–34	825 (23,6)	195 (24,4)	194 (24,7)	172 (26,0)
35–39	920 (26,3)	219 (27,3)	216 (27,5)	191 (28,9)
Total	3.494 (100)	800 (100)	785 (100)	661 (100)
<i>Nuuk (Groenlandia)</i>				
20–24	582 (37,0)	281 (35,1)	227 (32,7)	193 (32,9)
25–29	439 (27,9)	226 (28,3)	192 (27,6)	171 (29,2)
30–34	328 (20,8)	167 (20,8)	156 (22,4)	127 (21,7)
35–39	225 (14,3)	126 (15,8)	120 (17,3)	95 (16,2)
Total	1.574 (100)	800 (100)	695 (100)	586 (100)

^a Datos de Kjaer *et al.* (1988)

Tabla 10.3. Distribución de edad de las poblaciones femeninas residentes de Nykøbing Falster (Dinamarca) y Nuuk (Groenlandia), de las dos muestras aleatorias elegidas a partir de las mismas, y de las participantes en el estudio^a.

En el Ejemplo 10.11, la distribución de la edad de las participantes fue bastante similar a la de la población femenina total en las dos zonas elegidas, con una representación ligeramente menor de las mujeres en el grupo de edad más joven.

10.1.7 Comentarios finales

Los métodos para elegir una muestra adecuada constituyen un campo de la estadística importante y bien desarrollado y no pueden tratarse ampliamente en este capítulo. No obstante, cabe destacar que el diseño del muestreo debe ser adecuado para los objetivos específicos del estudio. Para desarrollar e implementar un esquema de muestreo adecuado no sólo es necesario tener experiencia en estadística sino también familiaridad con las condiciones del terreno. Por ejemplo, si el estudio utiliza 'familias' como una de sus unidades de muestreo, este término debe definirse claramente en el contexto de la población en la que se va a realizar el estudio. Una 'familia' se define normalmente como un grupo de personas que viven y comen juntos. No obstante, en muchas sociedades esta definición no se traducirá fácilmente a la práctica. En estos casos, es decisivo conocer bien y definir claramente las diferentes organizaciones de la población a estudiar antes de iniciar el estudio.

En la fase de diseño del estudio deben preverse otros problemas prácticos y debe proporcionarse a los entrevistadores instrucciones inequívocas, escritas en el protocolo. Por ejemplo, si no hay nadie en casa cuando llega el entrevistador, éste debe regresar de nuevo en lugar de ir a la casa del vecino, porque las familias con una persona en casa durante el día tienden a diferir de las que no la tienen.

Aunque el empleo de un método aleatorio para elegir la muestra del estudio es el método más apropiado, por razones éticas o logísticas, en gran cantidad de estudios epidemiológicos no es posible elegir una muestra de este modo. Por ejemplo, los estudios que requieren el uso de técnicas de diagnóstico agresivas que sólo pueden realizarse en un hospital, puede que se realicen únicamente con los pacientes que acuden al centro. Las conclusiones de estos estudios pueden extrapolarse a la población de los que 'acuden al hospital', pero la generalización a toda la población objeto de estudio requiere una evaluación cuidadosa y prudente.

La elección de un diseño de muestreo adecuado en epidemiología del cáncer también dependerá del objetivo principal del estudio. Si el objetivo principal es obtener una estimación global de la prevalencia (o nivel) de un atributo en la población objeto de estudio, deben utilizarse métodos de muestreo aleatorios en todas las fases del proceso para asegurar que no se introduzca un sesgo de selección. No obstante, si el objetivo principal del estudio es examinar las relaciones potenciales de dosis-respuesta, puede ser más adecuado elegir las unidades principales de muestreo de manera no aleatoria ya que, en estas situaciones, suele ser más importante la capacidad de informar que la representatividad. Por ejemplo, la selección de los 65 distritos chinos participantes en el Ejemplo 10.7 no se realizó aleatoriamente porque el objetivo principal del estudio no era proporcionar una estimación global de la prevalencia (o nivel) de los diferentes atributos del estilo de vida para toda China. El objetivo principal del estudio era comparar la distribución del estilo de vida en los distritos que se sabía presentaban unos niveles muy diferentes de mortalidad por algunos tipos de cáncer. No obstante, era necesario obtener unas estimaciones no sesgadas de la prevalencia del estilo de vida

para cada uno de los distritos elegidos y por esta razón se utilizaron métodos aleatorios para seleccionar a los participantes en cada distrito.

Otro aspecto a considerar en el diseño del muestreo es el tamaño de la muestra. La muestra debe ser lo suficientemente grande para tratar los objetivos principales del estudio con una precisión adecuada, pero no excesivamente mayor de lo necesario, de modo que no se malgasten innecesariamente los recursos. Los aspectos del tamaño de la muestra se comentarán en el Capítulo 15.

10.2 Obtención de datos

Los métodos utilizados para recopilar los datos pertinentes en estudios transversales son básicamente los comentados en el Capítulo 2. Los cuestionarios, a veces complementados con pruebas de diagnóstico y obtención de muestras biológicas, son la manera más habitual de obtener información de los sujetos. La mayoría de los cuestionarios incluye preguntas sobre exposiciones en el pasado así como exposiciones actuales. La información sobre exposiciones en el pasado aumenta considerablemente la capacidad de los estudios de identificar las relaciones dosis-respuesta.

10.3 Análisis

*Prevalencia*⁵ es la medida de aparición de una enfermedad, condición o característica que puede obtenerse principalmente a partir de estudios transversales (véase Apartado 4.2.1). Esto se ilustra en el Ejemplo 10.12.

Ejemplo 10.12. *Supongamos que se efectuó un estudio transversal para evaluar la prevalencia de los quistes de mama en una población femenina en particular. Se examinó una muestra de 5.891 mujeres elegidas al azar de esta población y un total de 201 presentaban quistes. La prevalencia de quistes de mama en esta población en el momento del estudio puede calcularse como: $201 / 5.891 = 3,4\%$.*

Para examinar la asociación entre un factor de riesgo para el atributo de interés, la población se subdivide en primer lugar en los expuestos y los no expuestos al factor en estudio y se calcula y compara la prevalencia del atributo en cada uno de estos dos grupos. A continuación puede calcularse una *razón de prevalencia* como la relación entre la prevalencia del atributo de interés en los expuestos al factor de riesgo respecto a la prevalencia en los no expuestos.

En el Ejemplo 10.13, la prevalencia de quistes de mama fue un 30% mayor en las usuarias de anticonceptivos orales en comparación con las que nunca los habían utilizado. Cabe destacar que la razón de prevalencias es una buena estimación de la razón de tasas de incidencia sólo si la prevalencia del resultado de interés entre las no expuestas es baja (inferior al 10%) y la duración de la enfermedad es la misma entre las expuestas y las no expuestas al factor de interés.

⁵ Cuando la característica de interés es una variable cuantitativa (como duración de la lactancia, peso, altura, etc.), la prevalencia tan sólo se puede calcular si las observaciones se clasifican en categorías. De lo contrario, se pueden utilizar medias o medianas.

Ejemplo 10.13. Supongamos que en el estudio hipotético descrito en el Ejemplo 10.12, los investigadores quieren evaluar si la prevalencia de quistes de mama estaba asociada con el uso de anticonceptivos orales. Los resultados se muestran en la Tabla 10.4.

Tabla 10.4. Quistes de mama y uso de anticonceptivos orales: datos de un estudio transversal hipotético.

Quistes de mama	Uso de anticonceptivos orales		Total
	Alguna vez	Nunca	
Sí	124	77	201
No	3.123	2.567	5.690
Total	3.247	2.644	5.891
Prevalencia de quistes de mama entre las usuarias = $124 / 3.247 = 3,8\%$			
Prevalencia de quistes de mama entre las no usuarias = $77 / 2.644 = 2,9\%$			
Razón de prevalencia = 1,3			

A menudo, las exposiciones de interés pueden clasificarse en varios niveles de intensidad, como muestra el Ejemplo 10.14. Podremos pues examinar tendencias en la prevalencia según el nivel de exposición.

También es habitual medir la fuerza de la asociación entre un factor de riesgo y el resultado de interés en un estudio transversal mediante el cálculo de la *odds ratio*. Ésta es la razón de la probabilidad de la exposición a un factor de riesgo entre aquellos con el resultado de interés y aquellos sin el resultado.

Mediante el cálculo de las *odds ratios*, el estudio transversal puede analizarse como si fuera un estudio caso-control. No obstante, los estudios transversales difieren de los estudios caso-control en que los 'casos' y los 'controles' se definen *a posteriori*, es decir, durante el análisis y no en la fase de diseño. De hecho, si el resultado de interés es cuantitativo, pueden efectuarse incluso varios análisis utilizando definiciones diferentes de 'casos' y 'controles' cambiando el punto de corte.

No obstante, cabe destacar que la *odds ratio* sólo proporcionará una buena estimación de la razón de prevalencias si la prevalencia de la enfermedad es baja (como en el Ejemplo 10.15).

En el Ejemplo 10.16, no sería adecuado tomar la *odds ratio* como una medida de la prevalencia relativa porque en este ejemplo la prevalencia de infección por VPH es relativamente alta (21,1% en mujeres que notificaron sólo una pareja).

10.4 Interpretación

Los estudios transversales son relativamente fáciles y económicos de realizar y son especialmente útiles para investigar exposiciones que son características fijas de los individuos, tales como grupo étnico y grupo sanguíneo.

No obstante, los estudios transversales no son el diseño de estudio adecuado para investigar las relaciones causales porque se basan en casos prevalentes en lugar de incidentes. Este tipo de estudios mostrará la presencia o ausencia de una relación entre las variables del estudio y los casos prevalen-

Ejemplo 10.14. Se realizó un estudio transversal con mujeres que acudieron a un centro sanitario universitario para investigar los determinantes de la infección en el cuello uterino por el virus del papiloma humano (VPH). Se solicitó a una muestra de 467 mujeres que cumplimentaran un cuestionario autoadministrado sobre variables sociodemográficas y de comportamiento sexual en el momento de su visita a la clínica. Para detectar la infección se utilizó el método de amplificación de ADN mediante reacción en cadena de la polimerasa. A continuación se examinó la prevalencia de la infección por VPH respecto al estado civil y el número de compañeros sexuales a lo largo de la vida (Ley et al., 1991). Los resultados se presentan en la Tabla 10.5.

	Nº de mujeres	% positivas al VPH	Razón de prevalencia (IC 95%)
Estado civil			
Solteras ^b	437	47,4	1,0
Casadas alguna vez	30	20,0	0,4 (0,2-0,9)
Nº de parejas sexuales a lo largo de la vida			
1 ^b	90	21,1	1,0 ^c
2-3	101	32,7	1,5 (0,9-2,4)
4-5	93	54,8	2,6 (1,7-4,0)
6-9	66	56,1	2,7 (1,7-4,3)
10+	102	68,6	3,3 (2,1-4,9)

^a Datos de Ley et al. (1991).

^b Tomada como categoría de referencia.

^c Test de χ^2 de tendencia = 53,10; 1 g.l.; $P < 0,0001$.

(Intervalos de confianza al 95% y test de χ^2 de tendencia calculados mediante las fórmulas indicadas en el Anexo 6.1.)

El cálculo de la razón de prevalencias para cada uno de los niveles de exposición se realizó formando tablas 2 x 2 como se muestra a continuación para mujeres con más de 10 parejas:

	Número de compañeros sexuales		Total
	10+	1	
Positivas al VPH	70	19	89
Negativas al VPH	32	71	103
Total	102	90	192

Prevalencia entre mujeres con 10+ parejas = $70/102 = 68,6\%$

Prevalencia entre mujeres con una pareja = $19/90 = 21,1\%$

Razón de prevalencias = $68,6\% / 21,1\% = 3,3$

Tabla 10.5. Prevalencia de la infección por VPH según el estado civil y el número de compañeros sexuales en la vida^a.

tes (existentes). Esto requiere interpretar los resultados con precaución, ya que los casos prevalentes pueden no ser representativos de todos los casos de la enfermedad. Los casos de corta duración, debidos a una recuperación o

Ejemplo 10.15. En el estudio hipotético presentado en el Ejemplo 10.13, podemos calcular la diferencia de haber utilizado alguna vez anticonceptivos orales entre mujeres con quistes de mama ('casos') y sin quistes de mama ('controles').

Odds de exposición a anticonceptivos orales entre 'casos' = $124/77 = 1,61$

Odds de exposición a anticonceptivos orales entre 'controles' = $3.123/2.567 = 1,22$

Odds ratio = $1,61 / 1,22 = 1,3$

Ejemplo 10.16. Utilizando los datos de la Tabla 10.6, podemos calcular la probabilidad de haber tenido 10 o más parejas ('exposición') entre las mujeres positivas al VPH ('casos') y las negativas al VPH ('controles') como sigue:

Odds de exposición entre los casos = $70/19 = 3,68$

Odds de exposición entre los controles = $32/71 = 0,45$

Odds ratio = $3,68/0,45 = 8,2$

Esta odds ratio de 8,2 se diferencia claramente de la razón de prevalencias de 3,3 calculada en el Ejemplo 10.14.

muerte rápidas, presentan una menor probabilidad de ser detectados en un estudio de prevalencia que los casos de mayor duración. Lógicamente los casos de larga duración están hiperrepresentados en un estudio transversal. Las características de estos casos de larga duración pueden diferir en muchos aspectos del conjunto de todos los casos de la enfermedad estudiada. Cuando algunos casos abandonan la comunidad de origen, los casos prevalentes pueden también convertirse en no representativos de todos los casos. Algunos sujetos afectados pueden institucionalizarse en otro lugar o trasladarse a otra ciudad en la que existan instalaciones especiales para el tratamiento.

Los estudios transversales también son una estrategia inadecuada para el estudio de enfermedades poco frecuentes ya que sería necesario estudiar una población muy grande para identificar un volumen suficiente de casos. Por tanto, su uso en epidemiología del cáncer se ha restringido a la investigación de factores asociados a lesiones precursoras.

Otra limitación principal de los estudios transversales es su dificultad de establecer la *secuencia* de los acontecimientos en el tiempo. Por ejemplo, en nuestro ejemplo hipotético de quistes de mama y el uso de anticonceptivos orales, no se puede presuponer que el uso de anticonceptivos orales precedió a los quistes. De hecho, a las mujeres con trastornos benignos de mama a veces se les prescribe anticonceptivos orales para mejorar su sintomatología. Por el contrario, no habría duda sobre la secuencia en el tiempo si el objeto de estudio fuera la presencia de un cáncer y rasgos como grupo sanguíneo o exposición de la madre a radiación.

Cuadro 10.1. Puntos esenciales

- Los *estudios transversales* son estudios en los que se selecciona a un grupo de sujetos (*muestra*) de una población definida (*población origen*) y se les entrevista en un momento determinado. Basándose en la información obtenida de los sujetos en este punto en el tiempo, se clasifican en función de si presentan o no el atributo de interés.
- Pueden utilizarse varios métodos para seleccionar una muestra representativa de la población origen. El *muestreo aleatorio* es el mejor porque asegura que sólo el azar determina quien se incluirá en la muestra, eliminando toda posibilidad de sesgo de selección. Pueden utilizarse diferentes diseños de muestreo para seleccionar una muestra aleatoria dependiendo de los objetivos específicos del estudio, la disponibilidad de un marco de muestreo adecuado, el tamaño y dispersión geográfica de la población de origen y los costes.
- La selección de una muestra aleatoria no elimina el sesgo de selección del estudio. El sesgo de selección todavía puede introducirse en el estudio si los que participan difieren de los que se niegan a participar o no se localizan. Por tanto, es importante asegurar un nivel de participación alto.
- Las principales ventajas de los estudios transversales son:
 - 1) Más fáciles de realizar que otros estudios debido a que no es necesario un seguimiento.
 - 2) Proporcionan una buena representación de las necesidades sanitarias de la población en un momento determinado.
 - 3) Pueden utilizarse para investigar exposiciones múltiples y resultados múltiples.
- Los principales inconvenientes de los estudios transversales son:
 - 1) Se basan en casos prevalentes (existentes) en lugar de incidentes (nuevos). Por tanto, presentan un valor limitado para investigar relaciones etiológicas.
 - 2) No son útiles para investigar enfermedades poco frecuentes o enfermedades de corta duración.
 - 3) No son adecuados para investigar exposiciones poco frecuentes.
 - 4) Es difícil establecer la secuencia de los acontecimientos en el tiempo.

Lecturas adicionales

- * El libro de Casley y Lury (1987) es un texto práctico sobre el papel, la planificación y la realización de estudios transversales en países en vías de desarrollo.
- * Ross y Vaughan (1986) proporcionan una revisión metodológica del uso de estudios transversales para obtener datos de morbilidad e información sobre el uso de servicios sanitarios en países en vías de desarrollo.

Capítulo 11

Estudios basados en datos rutinarios

Los estudios basados en datos rutinarios se caracterizan por el hecho de que los datos sobre la(s) exposición(es) y el(los) resultado(s) de interés se obtienen a partir de sistemas de recopilación de datos rutinarios (p.e. registros del cáncer, registros de hospitales, certificados de defunción, etc.). Por tanto, este tipo de estudio puede efectuarse de manera relativamente rápida y barata sin necesidad de ponerse en contacto con los sujetos del estudio. No obstante, su limitación principal es que el número de variables disponibles en los sistemas rutinarios de vigilancia epidemiológica generalmente es limitado.

Los estudios basados en datos rutinarios pueden llevarse a cabo a nivel *individual* o a nivel *agregado*.

11.1 Nivel individual

La mayoría de sistemas de recopilación de datos rutinarios reúnen datos sobre atributos personales tales como edad, sexo, lugar de nacimiento, lugar de residencia, ocupación, etc. La aparición de cáncer puede examinarse en relación a estas variables. El objetivo es buscar patrones que puedan sugerir o confirmar hipótesis etiológicas específicas.

11.1.1 Lugar de residencia

Estudiando la gran variabilidad en la incidencia del cáncer en el mundo, los estudios basados en datos rutinarios han proporcionado una indicación de la magnitud en que los factores ambientales están implicados en el origen de cada tipo de cáncer (Ejemplo 11.1).

El examen de las variaciones geográficas en la incidencia del cáncer puede proporcionar claves importantes para la etiología de esta enfermedad y estimular más investigaciones.

El cáncer de esófago es uno de los tipos de cáncer con los patrones más marcados de variación geográfica. Su incidencia varía mucho de una zona a otra, con áreas de riesgo muy alto completamente rodeadas por zonas de riesgo muy bajo. La variación geográfica en el riesgo de este cáncer en Asia Central estimuló la realización de una serie de estudios transversales y caso-control en el norte de Irán con el objetivo específico de investigar posibles factores de riesgo (Ejemplo 11.2). Estos estudios se interrumpieron en 1978 por la inestabilidad civil de la zona, pero los resultados iniciales indicaban una fuerte asociación geográfica con una dieta pobre en vegetales y con el uso de opio (*Joint Iran-International Agency for Research on Cancer Study Group, 1977*).

Ejemplo 11.1. Se utilizaron los datos de registros de cáncer de base poblacional de todo el mundo para examinar las variaciones internacionales en la incidencia de cáncer. Para la mayoría de los tipos de cáncer, había una variación de más de diez veces entre la tasa de incidencia más alta y la más baja registrada (Tabla 11.1).

Tabla 11.1.

Variaciones internacionales en la incidencia registrada para los tipos de cáncer seleccionados: la tasa más alta y más baja en todos los registros del cáncer de base poblacional incluidos en *Incidencia del Cáncer en Cinco Continentes Vol. V [Cancer Incidence in Five Continents, (CI5C)]*. Tasas por 100.000 personas-año, estandarizadas por edad para la población estándar mundial^a.

Tipo de cáncer (CIE 9)	Hombres			Mujeres		
	Tasa más alta (A)	Tasa más baja (B) ^b	Razón A:B	Tasa más alta (A)	Tasa más baja (B) ^b	Razón A:B
Labio (140)	15,1	0,1	151	1,6	0,1	16
Lengua (141)	9,4	0,4	24	3,4	0,2	17
Boca (143-145)	13,5	0,5	27	15,7	0,2	79
Nasofaringe (147)	30,0	0,3	100	12,9	0,1	129
Faringe (146, 148-149)	31,3	0,4	78	4,3	0,2	22
Esófago (150)	29,2	1,2	24	12,4	0,3	41
Estómago (151)	82,0	3,7	22	36,1	3,0	12
Colon (153)	34,1	1,8	19	29,0	1,8	16
Recto (154)	21,5	3,0	7	13,4	1,3	10
Hígado (155)	34,4	0,7	49	11,6	0,4	29
Laringe (161)	17,8	2,2	8	2,7	0,2	14
Pulmón (162)	110,0	5,8	19	68,1	1,2	57
Melanoma cutáneo (172)	30,9	0,2	155	28,5	0,2	143
Mama (174/175)	3,4	0,2	17	93,9	14,0	7
Cuello uterino (180)	—	—	—	83,2	3,0	28
Ovario (182)	—	—	—	25,8	3,3	8
Próstata (185)	91,2	1,3	70	—	—	—
Testículo (186)	8,3	0,6	14	—	—	—
Vejiga (188)	27,8	1,7	16	8,5	0,8	11
Sistema nervioso (191-192)	9,7	1,1	9	10,0	0,8	13
Linfoma no Hodgkin (200, 202)	11,4	1,5	8	8,7	0,9	10
Enfermedad de Hodgkin (201)	4,8	0,5	10	3,9	0,3	13

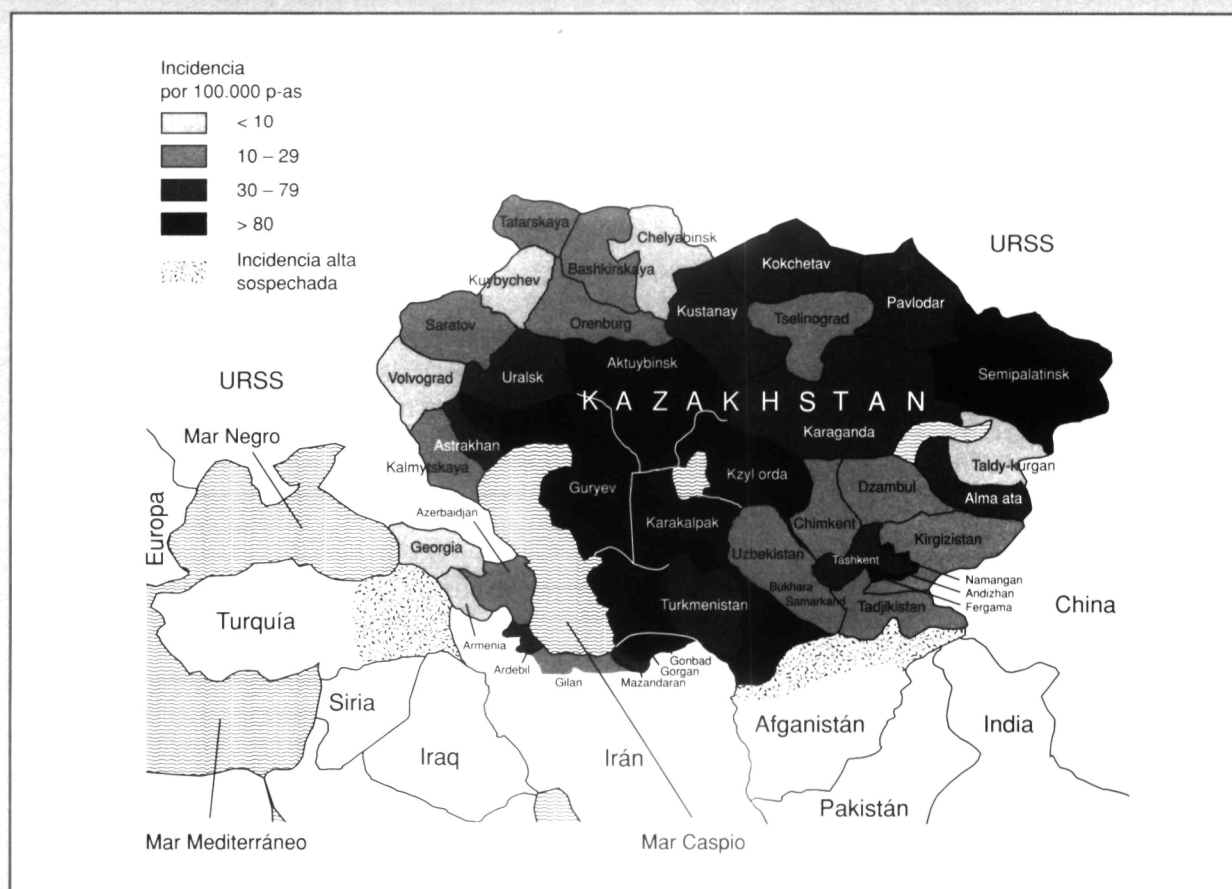
^a Datos de Whelan *et al.* (1990).

^b Se excluyeron las tasas basadas en menos de 10 casos.

Parte de las diferencias internacionales en los riesgos de cáncer pueden ser genéticas más que ambientales. Pero para muchos tipos de cáncer todavía existen variaciones marcadas en el riesgo entre países incluso cuando sus poblaciones son bastante homogéneas genéticamente.

La distribución mostrada en el Ejemplo 11.3 inicialmente se supuso que era debida al consumo de sidra de manzana casera. Normandía y Bretaña son

Ejemplo 11.2. Se utilizaron datos de registros del cáncer locales para examinar la distribución geográfica del cáncer de esófago en Asia Central. Los resultados de este análisis mostraron que la incidencia de este cáncer era extremadamente alta en una zona que incluye Kazajistán, Uzbekistán y Turkmenistán, el nordeste de Irán y el norte de Afganistán (Figura 11.1)



las únicas provincias en las que se produce sidra de manzana en cantidades considerables y esta bebida se consume mucho en las mismas provincias (Tuyns *et al.*, 1983). Investigaciones posteriores demostraron que todos los tipos de bebidas alcohólicas parecen aumentar el riesgo de cáncer de esófago.

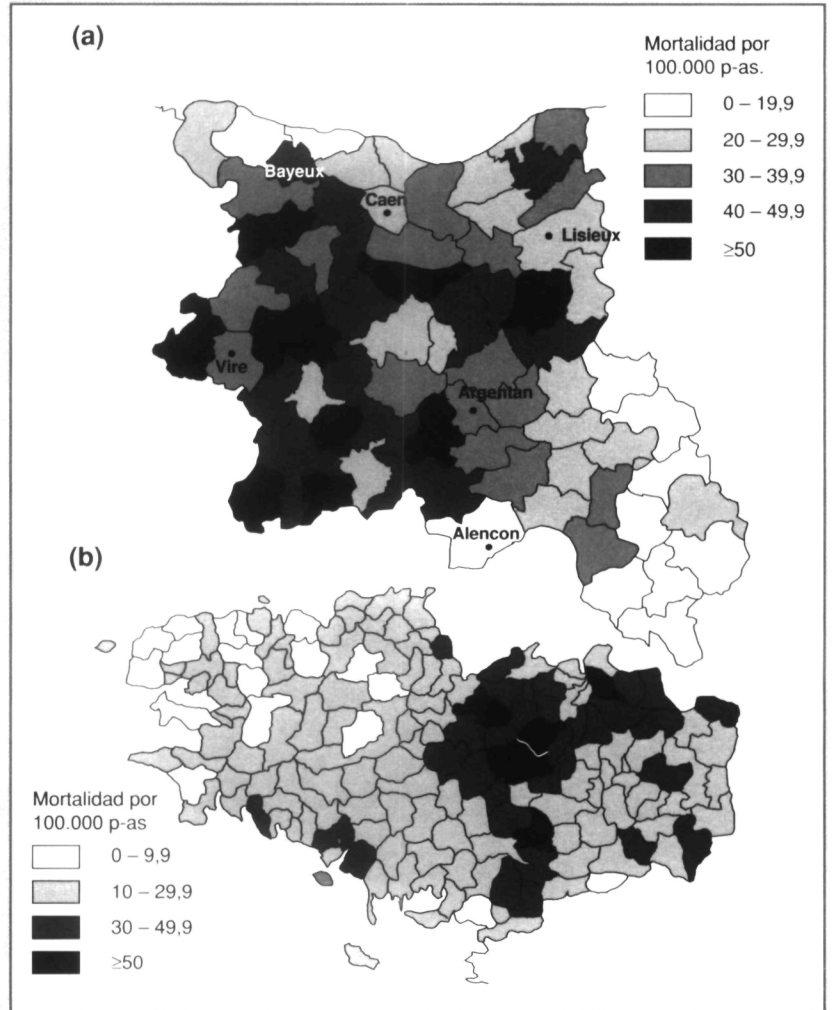
11.1.2 Lugar de nacimiento y grupo étnico

Las personas que migran de un país a otro presentan unas características del estilo de vida que son una combinación de las del país huésped y las de su país de origen. Por tanto, la evidencia de un gradiente de aumento o disminución de los riesgos entre la población de origen, los emigrantes y la población huésped, puede indicar o confirmar la importancia de los factores ambientales sobre los factores genéticos en la etiología de un cáncer en particular.

Figura 11.1. Cáncer de esófago en hombres de Asia Central. Tasas de incidencia estandarizadas por edad para la población estándar mundial (reproducido con la autorización de Oxford University Press, de Muñoz y Day, 1996).

Ejemplo 11.3. Los datos de mortalidad obtenidos rutinariamente se utilizaron para examinar la distribución geográfica del cáncer de esófago en la Bretaña y en los departamentos de Calvados y Orne en Normandía (Figura 11.2). Aunque las tasas en la mayoría de los cantones eran similares a la tasa media en Francia, en algunos cantones en el este de Bretaña y el noroeste de Orne la mortalidad fue de cinco a diez veces superior (Tuyns y Vernhes, 1981).

Figura 11.2. Tasas de mortalidad estandarizadas por edad para el cáncer de esófago en hombres, por cantón en (a) departamentos de Calvados y Orne en Normandía (reproducido con la autorización de Tuyns y Vernhes, 1981. © Masson Editeur, 1981) y (b) Bretaña, 1958-66 (modificado de Tuyns y Mas-sé, 1973).



Esta estrategia puede refinarse añadiendo un componente de tiempo. El grado de integración cultural en el país huésped puede medirse indirectamente por la información sobre la edad en el momento de la migración y/o el tiempo desde la migración.

En el Ejemplo 11.4, los inmigrantes europeos presentaban un riesgo menor de morir de melanoma que los nacidos en Australia, especialmente los

Ejemplo 11.4. Los datos sobre las muertes registrados en Australia durante el periodo 1964-1985 se obtuvieron de la Oficina Australiana de Estadística para examinar la mortalidad por melanoma maligno cutáneo en inmigrantes en comparación con individuos nacidos en Australia y para investigar los cambios en el riesgo con la edad de llegada y la duración de la permanencia. Cada registro de muerte contenía información sobre los siguientes aspectos: sexo, país de nacimiento, duración de la permanencia en Australia, año de fallecimiento, edad de fallecimiento y causa de muerte (Khalat et al., 1992). Algunos de los resultados de este estudio se presentan en la Figura 11.3.

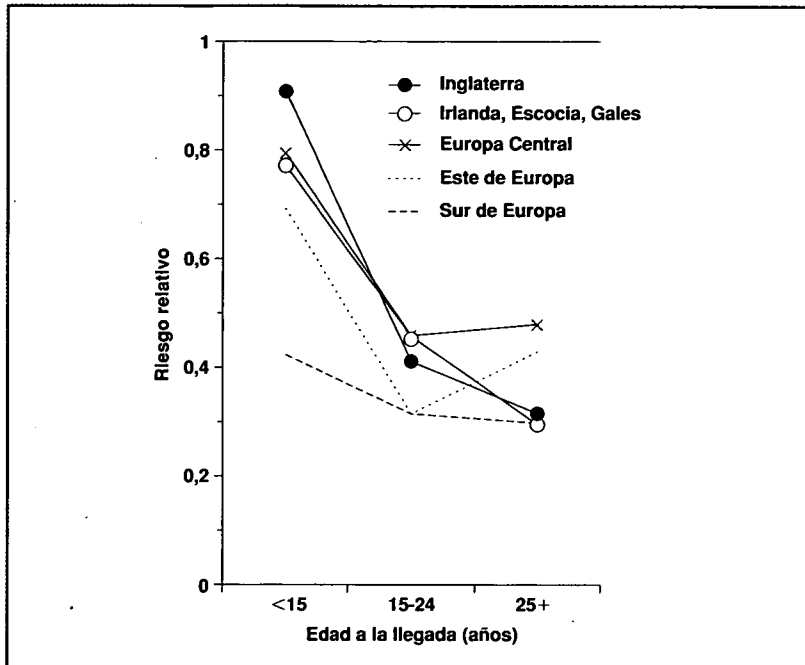


Figura 11.3.

Riesgos relativos (RR) de mortalidad por melanoma en inmigrantes europeos varones en Australia, por región de nacimiento y edad a la llegada, en comparación con los nacidos en Australia (considerado como el grupo basal: RR=1); Australia, 1964-1985 (reproducido con la autorización de Khalat et al., 1992).

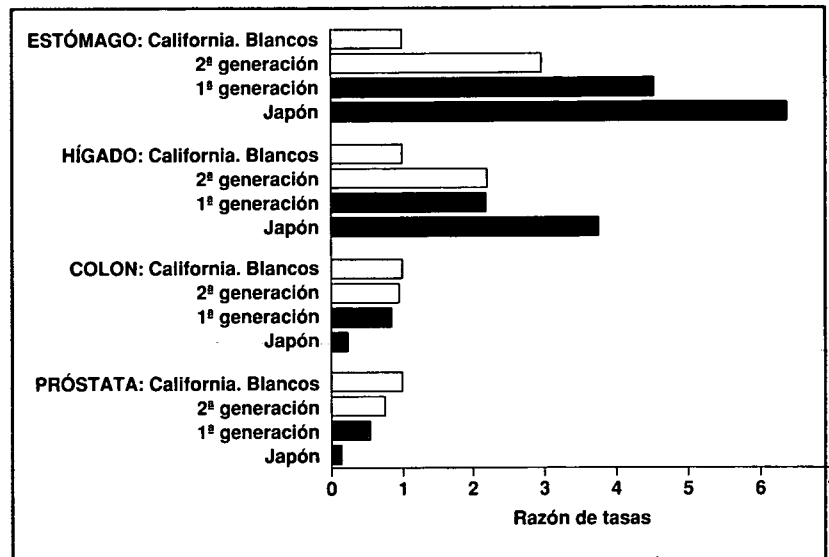
del sur de Europa. Esta observación puede reflejar el efecto protector de los caracteres más oscuros. Los datos también indicaban que la migración en la infancia estaba asociada a un riesgo mayor que la migración a una edad más avanzada pero es difícil separar este efecto del de la permanencia (es decir, la migración a una edad joven estaba asociada inevitablemente con una permanencia más prolongada que la migración a una edad más avanzada).

Si se dispone de la información sobre el lugar de nacimiento y la etnia puede distinguirse a los inmigrantes de primera generación (nacidos en el país de origen) respecto a sus hijos, con frecuencia nacidos en el país huésped, que se consideran inmigrantes de segunda generación. Esta distinción proporciona otro indicador útil del posible grado de asimilación de los inmigrantes al estilo de vida característico del país huésped, que tiende a ser mayor en la segunda que en la primera generación.

Ejemplo 11.5. En un estudio de inmigrantes japoneses residentes en California, se comparó la mortalidad por formas comunes de cáncer en los inmigrantes de primera y segunda generación con las tasas correspondientes para California y Japón (Buell y Dunn, 1965). La Figura 11.4 muestra los resultados de cuatro de estos tipos de cáncer más comunes: estómago, hígado, colon y próstata.

Figura 11.4.

Razones de tasas de mortalidad (RT) ajustadas por edad por cáncer de estómago, hígado, colon y próstata en hombres japoneses en Japón, e inmigrantes japoneses a California de primera y segunda generación, comparados con hombres blancos de California (considerados como el grupo de referencia: RT=1); Japón, 1952-59 y California 1956-62 (datos de Buell y Dunn, 1965).



En el Ejemplo 11.5 había un patrón coincidente: el riesgo para todos estos tipos de cáncer convergía hacia el riesgo en el país huésped. Para los cánceres de estómago e hígado, ambos eran más frecuentes en Japón que entre los blancos de los Estados Unidos; en cambio, los hombres nacidos en Japón pero emigrantes a California presentaban unos riesgos de muerte considerablemente menores que los hombres de la misma edad en Japón, sobre todo los riesgos en la segunda generación. Por el contrario, el riesgo de los cánceres de colon y próstata eran más frecuentes en California que en Japón. En los emigrantes el riesgo era similar al de los hombres blancos de California. Estas observaciones muestran claramente la importancia de los factores ambientales sobre los factores genéticos relacionados con el grupo étnico. Todavía se observaba una diferencia residual entre tasas de los inmigrantes de segunda generación y las tasas de California. Estas diferencias residuales pueden reflejar diferencias en la susceptibilidad genética a estas enfermedades o, de manera alternativa, pueden indicar que los inmigrantes de segunda generación conservaban todavía algunos de los comportamientos «tradicionales» del estilo de vida japonés.

No obstante, cualquier diferencia en los riesgos de cáncer entre emigrantes y aquellos que permanecieron en su país de origen debe interpretarse con precaución, debido a que los emigrantes suelen ser un grupo no repre-

Ejemplo 11.6. La mortalidad por cáncer de páncreas en hombres emigrantes italianos a ocho países huéspedes se comparó con la mortalidad en su país de nacimiento (Italia). La mayoría de los emigrantes italianos procedían del sur de Italia. Por ello, se intentó evitar un sesgo de selección examinando al mismo tiempo los riesgos relativos con los del sur de Italia y con los de todo el país. La Figura 11.5 muestra, por ejemplo, que la mortalidad por cáncer de páncreas era mucho menor en el sur de Italia (Riesgo relativo (RR) = 0,38) y en toda Italia (RR = 0,65) que en Canadá (considerado el valor basal: RR=1). La mortalidad en los emigrantes italianos a Canadá era próxima a la de la población huésped (RR = 0,93) (Balzi et al., 1993).

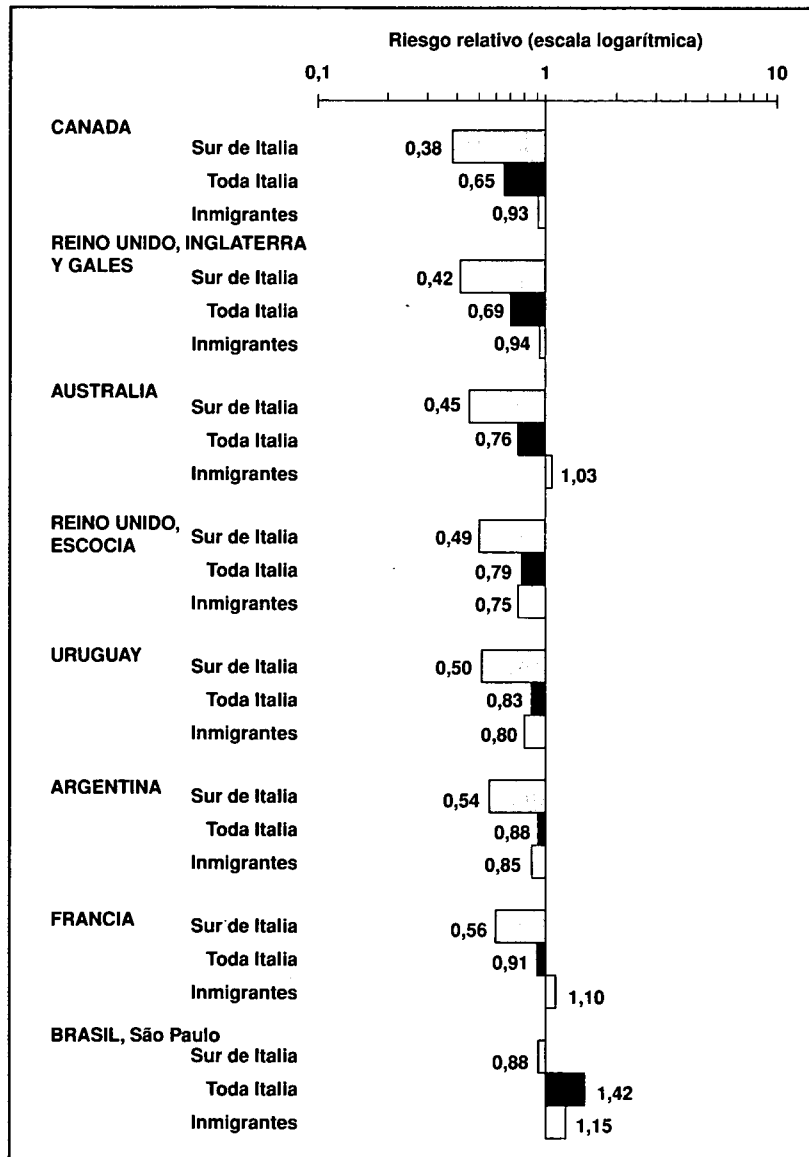


Figura 11.5. Riesgo relativo (RR) de mortalidad por cáncer de páncreas en Italia (en el sur y en toda Italia), y en ocho poblaciones masculinas italianas inmigrantes comparado con las poblaciones de los ocho países huéspedes (cada población huésped es considerada como el grupo de referencia: RR=1) (reproducido con la autorización de Balzi et al., 1993).

sentativo de las poblaciones de su país de origen. También es probable que los emigrantes difieran de la población huésped en una serie de características demográficas y socioeconómicas, lo que debe tenerse en cuenta al comparar los riesgos.

En el Ejemplo 11.6, la mayoría de los emigrantes procedían del sur de Italia, que presentaba una mortalidad por cáncer de páncreas menor que el país en conjunto.

En estudios basados en datos de rutina, los emigrantes se identifican en las series de datos rutinarios basándose en el lugar/país de nacimiento, la raza y, cuando no se dispone de información sobre ninguna de estas dos variables, basándose en el análisis del nombre. Más raramente se ha utilizado información sobre la lengua y la religión. No obstante, cabe considerar que estas aproximaciones no proporcionan resultados similares. Por ejemplo, una proporción de personas nacidas en el subcontinente indio que emigraron a Inglaterra y Gales, de hecho, son de raza blanca. Así, el análisis basado exclusivamente en el país de nacimiento puede incluir emigrantes de primera generación independientemente de su raza. Por el contrario, el análisis basado exclusivamente en la raza incluirá emigrantes de un grupo étnico particular independientemente de si son de la primera generación o de generaciones subsiguientes.

11.1.3 Nivel socioeconómico y ocupación

Los análisis del riesgo de cáncer por grupo socioeconómico y ocupacional también han proporcionado ideas importantes en la epidemiología del cáncer.

El Ejemplo 11.7 muestra unas asociaciones importantes entre la mortalidad y el nivel de educación. El gradiente fue positivo para los cánceres de mama y de colon pero negativo para los cánceres de estómago y cuello uterino (Tabla 11.2). Estos gradientes son similares a los observados en los países occidentales.

En el Ejemplo 11.8, los hombres con ocupaciones al aire libre presentaban la mayor incidencia proporcional de cáncer de labio (Tabla 11.3). Estas observaciones parecen indicar que la exposición al sol puede ser importante en la etiología de este cáncer.

Las personas con ocupaciones diferentes tienden a tener estilos de vida diferentes y exposiciones ocupacionales diferentes. De este modo, las diferencias en sus riesgos de cáncer pueden proporcionar también claves etiológicas respecto a factores que no están relacionados directamente con la propia ocupación (como en el Ejemplo 11.9).

La Tabla 11.4 muestra un gradiente por clase social importante en el riesgo de muerte por cáncer de cuello uterino en las mujeres casadas de Inglaterra y Gales. Las esposas de hombres de la clase social I (clase social más alta) experimentaron una tasa de mortalidad de sólo el 34% de la tasa para todas las mujeres casadas. Por el contrario, la mortalidad de las esposas de hombres de la clase social V (clase social más baja) fue del 81% mayor que la de todas las mujeres casadas. No obstante, dentro de cada clase social, las tasas más altas eran las de las esposas de hombres con ocupaciones que im-

Ejemplo 11.7. Los registros de defunción de los residentes en São Paulo (Brasil) durante los años 1978-82 se examinaron para investigar los gradientes socioeconómicos en la mortalidad por cáncer. Los certificados de defunción contenían información sobre el sexo, edad, lugar de nacimiento y residencia, causa de muerte, estado civil y última ocupación del fallecido. La familia del fallecido proporcionó información sobre la educación, indicada por los años de escolarización, recogidos en cinco categorías: menos de 1, 1 a 8, 9 a 11, 12 ó más, y desconocido (Bouchardy et al., 1993). La Tabla 11.2 muestra los diferenciales de educación en la mortalidad por un grupo seleccionado de tipos de cáncer femeninos.

Tipo de cáncer (Código CIE-9)	Nº de muertes	Educación desconocida (%)	Odds ratio por años de educación ^b				Test de χ^2 de tendencia ^c
			<1 ^d	1-8	9-11	>11	
Estómago (151)	691	6,2	1,0	0,7	0,3	0,3	43,5; $P < 0,001$
Colon (153)	338	7,1	1,0	1,4	2,1	2,2	13,3; $P < 0,001$
Mama (174)	1.744	5,9	1,0	1,6	2,4	2,6	77,5; $P < 0,001$
Cuello uterino (180)	645	7,2	1,0	0,7	0,4	0,2	50,2; $P < 0,001$

^a Datos de Bouchardy et al. (1993)

^b Los controles son todos los cánceres excepto el estudiado (veáanse los comentarios del método de análisis en el Apartado 11.1.6). Las odds ratios se ajustan por edad, estado civil y país de nacimiento.

^c Tras la exclusión de sujetos con nivel de educación desconocido

^d Considerado como la categoría de referencia

Tabla 11.2.

Mortalidad por tipos de cáncer seleccionados en relación a los años de escolarización. Mujeres de 35-64 años, São Paulo (Brasil), 1978-82^a.

Ejemplo 11.8. Los registros de cáncer en Inglaterra y Gales recopilan datos sobre la ocupación de los casos de cáncer en el momento de su diagnóstico. Estos datos se examinaron para identificar los grupos ocupacionales asociados con una alta incidencia de cáncer de labio (OPCS, 1978). Los resultados se muestran en la Tabla 11.3.

Ocupación	Número de casos observados	Razón de incidencia proporcional (RIP) ^b
Labradores, granjeros jardineros	23	2,18
Agricultores	35	5,26
Albañiles	14	2,39
Trabajadores de la construcción	16	2,35

^a Datos de OPCS (1978)

^b Ajustada por edad $P < 0,01$. Para todas las RIP mostradas en la tabla.

(Véase en el ejemplo 11.14 una ilustración del cálculo de las razones de incidencia proporcionales).

Tabla 11.3.

Incidencia de cáncer labial en hombres de 15-74, Inglaterra y Gales, 1968-69, por ocupación: ocupaciones asociadas con unas razones de incidencia proporcionales aumentadas y estadísticamente significativas^a.

Tabla 11.4.

Mortalidad por cáncer de cuello uterino entre mujeres casadas en función de la clase social y ocupación, Inglaterra y Gales, 1959-63^a. (En la tabla sólo se muestran los datos de las ocupaciones con los niveles de mortalidad más altos y más bajos para ilustrar el intervalo de riesgos dentro de cada clase social)

Ejemplo 11.9. Los datos de mortalidad recopilados rutinariamente para Inglaterra y Gales se utilizaron para examinar los riesgos de mortalidad por cáncer de cuello uterino entre mujeres casadas en función de la ocupación del marido y clase social (Beral, 1974). (En este país, los registradores de la mortalidad deben introducir la ocupación del marido en los certificados de defunción de las mujeres casadas y viudas; la clase social de mujeres casadas se determina a partir de la ocupación del marido). Los resultados se presentan en la Tabla 11.4.

Clase social	Ocupación del marido	Razón de mortalidad estandarizada (RME%) ^b
I	Todas las ocupaciones	34
	Sacerdotes	12
	Científicos	17
	Ingenieros civiles	60
II	Todas las ocupaciones	64
	Profesores	30
	Funcionarios	40
	Mesoneros	120
	Hosteleros	150
III	Todas las ocupaciones	100
	Peones de obras	40
	Administrativos	64
	Operadores de grúas	159
	Camioneros	168
IV	Todas las ocupaciones	116
	Comerciantes y ayudantes	71
	Jardineros y labradores	98
	Pescadores	257
	Clasificación de puentes e ingeniería	263
	Barqueros y marinos	
V	Todas las ocupaciones	181
	Limpiadores de ventanas y oficinas	95
	Jornaleros	222

^a Datos de Beral (1974)

^b Las tasas específicas por edad para todas las mujeres casadas en Inglaterra y Gales se toman como el estándar.

plican viajes y ausencia del hogar durante periodos prolongados, una situación que se sabe está asociada con un alto riesgo de enfermedades venéreas entre hombres. Este patrón avaló la hipótesis de que el cáncer de cuello uterino puede ser una infección transmitida sexualmente mucho antes de la identificación de un agente infeccioso (Beral, 1974).

En la interpretación de diferencias en la incidencia (y mortalidad) entre grupos socioeconómicos, debe recordarse que la propia salud puede deter-

minar la inclusión en un grupo ocupacional específico. Por ejemplo, las personas con mala salud, normalmente están obligadas a trabajar en ocupaciones que requieran menos esfuerzo físico y las ocupaciones que requieren más esfuerzo físico incluyen selectivamente sólo a personas con una buena salud.

11.1.4 Tendencias temporales

La información sobre cambios en el riesgo de cáncer a lo largo del tiempo puede generar hipótesis etiológicas o confirmar asociaciones sospechadas entre factores de riesgo y enfermedad. Además, mientras que la existencia de una variación geográfica en la incidencia entre poblaciones puede explicarse por diferencias genéticas, los cambios en la incidencia en una misma población implican mucho más claramente la introducción o desaparición de factores de riesgo ambientales.

Ejemplo 11.10. Los datos de todos los casos de cáncer de tiroides incidentes entre 1935-75 se extrajeron del Registro de Tumores de Connecticut (EE.UU.) para examinar los cambios temporales en las tasas de este cáncer y sus cuatro tipos histológicos principales (Pottern et al., 1980). Los cambios temporales de carcinoma papilar, uno de los cuatro tipos histológicos, se muestran en la Figura 11.6.

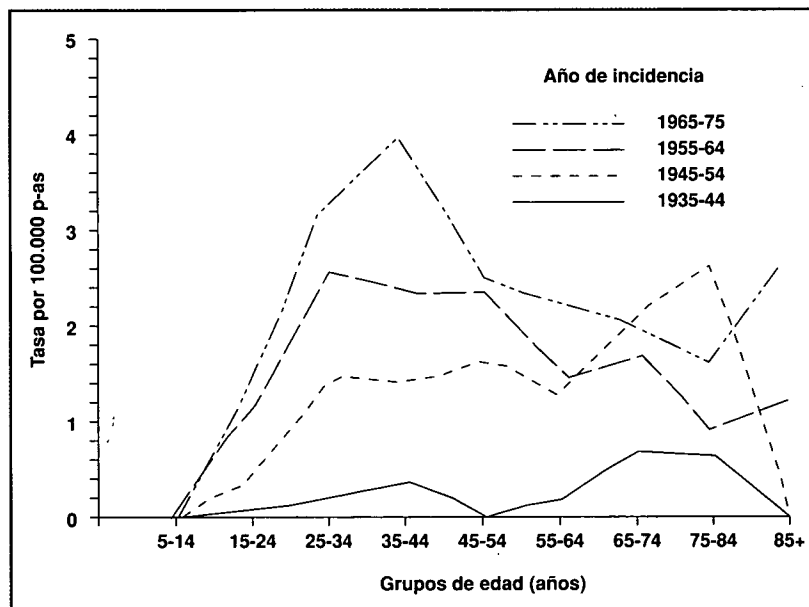


Figura 11.6.

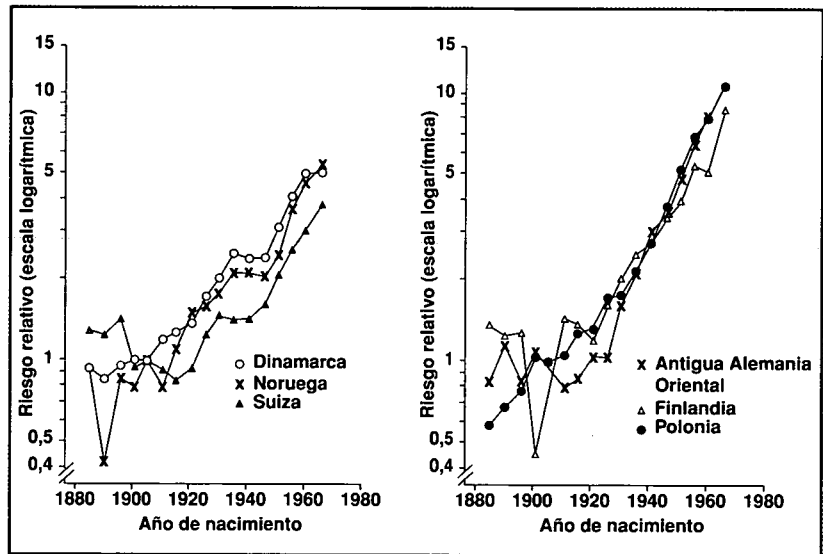
Tasas de incidencia específicas por edad para el carcinoma papilar de tiroides en mujeres por año de incidencia. Connecticut (EE.UU.) 1935-75 (reproducido con la autorización de Pottern et al., 1980).

En el Ejemplo 11.10 se muestra que la incidencia de carcinoma papilar aumentó de manera constante en los grupos de edad más jóvenes durante los 40 años del periodo de estudio, presentando un máximo en las edades de 25-44 años (Figura 11.6). Este aumento repentino en la incidencia en los

Figura 11.7.

Riesgo relativo (RR) de desarrollar cáncer de testículo por país y año de nacimiento. (Los hombres nacidos entre 1900 y 1909 se tomaron como la categoría de referencia: RR=1) (reproducido con la autorización de Bergström *et al.*, 1996).

Ejemplo 11.11. Se obtuvieron los datos de seis países de Europa (Dinamarca, Finlandia, la antigua República Democrática Alemana, Noruega, Polonia y Suecia) en los cuales existe un registro del cáncer desde los años 40 a 50, para examinar las tendencias a largo plazo de la incidencia de cáncer de testículo. Entre todos abarcan una población de aproximadamente 76 millones de personas. Se identificaron un total de 30.908 casos incidentes de cáncer de testículo diagnosticados durante 1945-84 en hombres de 20-84 años (Bergström *et al.*, 1996). La Figura 11.7 muestra las tendencias de cohorte (es decir tendencias por año de nacimiento de los hombres) en cada país.



EE.UU. se produjo después del uso generalizado de la radioterapia para tratar el 'timo aumentado' y otras patologías benignas de la cabeza y cuello en niños y adolescentes entre principios de los años 20 y finales de los 50 (Pottern *et al.*, 1980). No se observaron aumentos similares en la incidencia de este cáncer en poblaciones en las que esta terapia no había sido frecuente. Este ejemplo proporciona una buena ilustración de un aumento importante de la incidencia de un cáncer en particular debido a la introducción de una exposición específica.

En el Ejemplo 11.11 no se observó un cambio importante en el riesgo de cáncer de testículo para las generaciones sucesivas de hombres nacidos entre 1880 y 1920. No obstante, para las generaciones nacidas más adelante, el riesgo aumentó de manera constante excepto en los hombres nacidos durante 1930-45 en Dinamarca, Noruega y Suecia (los años alrededor de la Segunda Guerra Mundial). No se conocen las razones del marcado aumento de la incidencia. El aumento se ha producido principalmente en países industrializados y en grupos socioeconómicos altos, indicando que los factores asociados a estilos de vida opulentos pueden ser responsables. Paradó-

ticamente, la ligera disminución del riesgo en generaciones nacidas durante la Segunda Guerra Mundial se produjo en Dinamarca, Noruega o Suecia, que estuvieron mucho menos afectados por la guerra que Polonia, la antigua Alemania Oriental y Finlandia.

Las tendencias en el tiempo también son útiles para evaluar y monitorizar la eficacia de las actividades de control del cáncer tales como programas generalizados de cribaje. En el Ejemplo 11.12, se observó una relación estrecha entre la disminución de la incidencia en cada país y la cobertura ofrecida (o no ofrecida) por sus programas generalizados de cribaje organizado. La disminución de la incidencia fue más marcada en Finlandia e Islandia, países en los que los programas nacionales de cribaje se iniciaron a principios de los años 60. La disminución fue menos marcada en Suecia, donde el programa se introdujo gradualmente y en Dinamarca donde sólo un 40% de la población vivía en zonas con cribaje generalizado organizado. En Noruega, el único país que no tenía un programa organizado (excepto en una región), no se observó una disminución evidente de la incidencia de cáncer de cuello uterino.

Ejemplo 11.12. Se extrajeron los datos de seis registros nacionales nórdicos de cáncer para examinar las tendencias en el tiempo en la incidencia de cáncer de cuello uterino en cada uno de los países respecto a sus actividades de cribaje (Hakama, 1982).

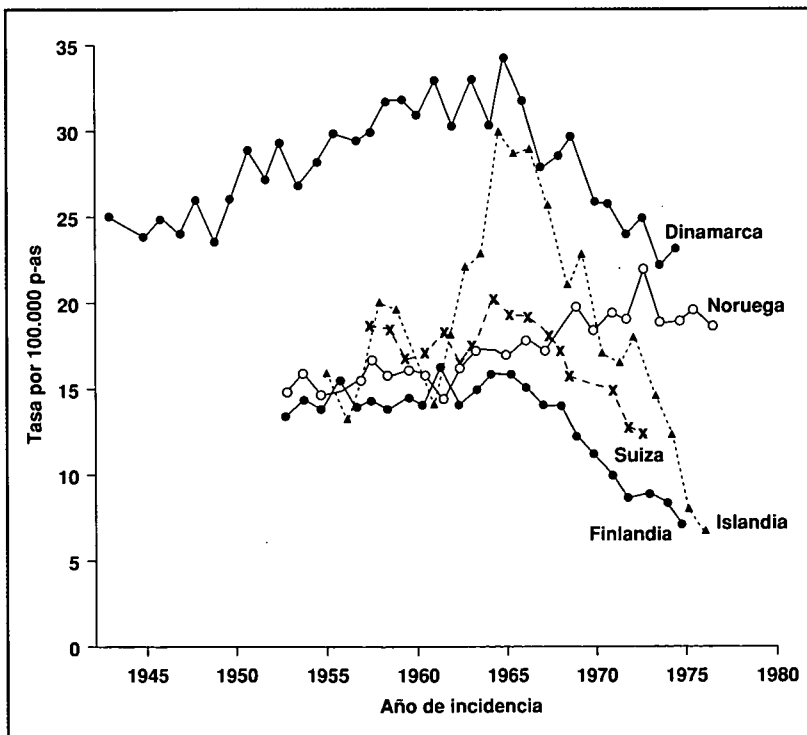


Figura 11.8. Incidencia anual de cáncer de cuello uterino en los seis países nórdicos, 1943-1978 (reproducido con la autorización de Hakama (1982)).

11.1.5 Estudios basados en el cruzamiento de registros

El cruzamiento de los datos de los registros del cáncer con archivos de otras fuentes tales como los datos del censo, los datos de mortalidad, archivos ocupacionales, archivos de admisión en hospitales, etc., se ha llevado a cabo para investigar los factores de riesgo de un gran número de cánceres tales como los cánceres relacionados con la ocupación y con la reproducción (véase Apartado 2.9.3).

Ejemplo 11.13. *En Dinamarca, se realizó un estudio para evaluar si la vasectomía podía aumentar el riesgo de cánceres de testículo y próstata. El estudio se basó en un cruce informatizado entre cuatro registros de base poblacional. El cruzamiento se realizó utilizando el número de identificación personal único asignado a cada residente de Dinamarca.*

El registro danés de altas hospitalarias ha archivado todas las admisiones hospitalarias en el país desde 1977, por fechas de admisión y alta, diagnóstico y operaciones. No obstante, algunos departamentos hospitalarios consideraron la vasectomía como un servicio ambulatorio y no lo notificaron al registro.

Los registros de anatomía patológica se establecieron en algunos distritos para archivar todas las muestras analizadas en los departamentos de anatomía patológica de hospitales dentro de su distrito específico. Como procedimiento rutinario se envía el tejido extraído en la vasectomía para un examen anatómico-patológico. De este modo, en los distritos en los que existen registros de anatomía patológica, todos los hombres vasectomizados (tanto los tratados como pacientes ingresados como los ambulatorios) pueden identificarse a partir de sus archivos.

El registro central de población danés es un archivo informatizado de todos los habitantes vivos en 1968 más los nacidos en Dinamarca o los inmigrantes a partir de este año, e incluye información sobre el estado vital y la fecha de emigración o muerte.

El registro del cáncer danés incluye todos los casos de cáncer diagnosticados en Dinamarca desde 1943, basándose en las notificaciones de los departamentos de hospitales, especialistas, informes de autopsias y certificados de defunción.

En el presente estudio, se estudiaron las altas hospitalarias y los registros de anatomía patológica para identificar a todos los hombres sometidos a vasectomía desde 1977 a 1989. Se identificaron un total de 73.917 hombres y a continuación se cruzaron sus registros con el registro central de población para obtener información sobre su estado vital el 31 de diciembre de 1989 y las fechas de emigración o de muerte. La aparición de cáncer en los miembros de la cohorte se identificó por el cruce de archivos con el registro del cáncer danés (Møller et al., 1994).

En el Ejemplo 11.13, a partir de altas hospitalarias y registros de anatomía patológica en Dinamarca se identificó retrospectivamente una cohorte

de 73.917 hombres que habían sido sometidos a vasectomía entre los años 1977-89. Se siguieron pasivamente desde el momento de la operación hasta el 31 de diciembre de 1989, cuando se evaluó su estado vital y de migración por el cruce con el registro de la población. La aparición de cáncer en miembros de la cohorte se determinó por el cruce con el registro nacional del cáncer.

11.1.6 Análisis

Los estudios basados en datos rutinarios realizados a nivel individual pueden considerarse estudios de cohorte en los que un grupo de personas, o cohorte, se sigue en el tiempo. Por ejemplo, la comparación de la incidencia o la mortalidad por cáncer en diferentes grupos ocupacionales puede considerarse como la comparación de la experiencia de cáncer entre grupos de personas con 'exposiciones' ocupacionales diferentes que se siguieron en el tiempo. Por tanto, el análisis de estos estudios es similar al análisis de cualquier otro estudio de cohorte y los métodos son básicamente los descritos en los Capítulos 4 y 8. El análisis se basa en el cálculo de las tasas como medidas de aparición de la enfermedad y de las *razones de tasas y diferencias de tasas* como medidas del efecto. El análisis de las tendencias en el tiempo por periodo de calendario y cohorte de nacimiento se consideró en el Apartado 4.3.2.

El cálculo de las tasas requiere información sobre la magnitud *persona-tiempo a riesgo* (es decir, denominadores) para cada uno de los grupos de interés. No obstante, no siempre se dispone de las estimaciones de la población que permitan estos cálculos. Por ejemplo, aunque la oficina nacional de estadística de muchos países puede proporcionar estimaciones de la población de todo el país y de zonas geográficas relativamente grandes, es posible que no sean capaces de proporcionar datos de zonas geográficas más pequeñas.

Además, los registros de cáncer y los certificados de defunción pueden contener información sobre muchas variables para las cuales no se obtienen datos en el censo (o cualquier otro sistema de recuento de la población). Por ejemplo, los registros del cáncer pueden recopilar información sobre el grupo étnico, pero si la información sobre esta variable no se obtiene en el censo, es posible que no se disponga de unos denominadores adecuados. En otras situaciones, los denominadores están disponibles en el censo, pero la información sobre los casos es tan incompleta que los datos del numerador no son comparables con los datos disponibles del denominador (por ejemplo, si en una gran proporción de los certificados de defunción faltan los datos sobre la ocupación del fallecido).

La opción tradicional cuando no se dispone de denominadores adecuados es calcular las *razones de incidencia o mortalidad proporcional* (ver Apartado 4.3.5). Esto se ilustra en el Ejemplo 11.14.

Las razones de incidencia (o mortalidad) proporcional suelen estar ajustadas por edad mediante el método indirecto de estandarización de un modo similar al utilizado en el cálculo de las razones de incidencia (o mor-

Ejemplo 11.14. Supongamos que un grupo de investigadores estuviera interesado en examinar la incidencia de cáncer de pulmón en hombres que trabajan en la industria de artes gráficas. La información sobre la ocupación de todos los casos nuevos de cáncer de pulmón que se produjeron en una región particular A durante los años 1970-74 estaba disponible a partir del registro local del cáncer de base poblacional, pero no se disponía de las estimaciones de la población por grupo de ocupación a partir del censo. En este caso, no fue posible calcular las tasas ya que no se disponía de los datos del denominador. En su lugar, se calcularon las medidas de la incidencia proporcional.

Supongamos que se produjeron un total de 10.000 cánceres incidentes en hombres en la región A durante los años 1970-74, de los cuales 2.000 eran cánceres de pulmón. Así, la proporción de cánceres de pulmón masculino en esta región fue

$$2.000 / 10.000 = 0,20$$

Supongamos también que el número total de cánceres incidentes en hombres impresores durante el mismo periodo (1970-74) fue de 100, de los cuales 40 eran cánceres de pulmón. La proporción de cánceres de pulmón masculino entre los impresores fue

$$40 / 100 = 0,40$$

Podemos calcular el número de casos de cáncer de pulmón que se esperaría (E) en los impresores si presentaran la misma proporción de cánceres de pulmón que la población total de la región. Esto sería igual a

$$E = 100 \times 0,20 = 20$$

La razón de incidencia proporcional (RIP) puede obtenerse por división del número observado de cánceres de pulmón en los impresores (O) por el número esperado (E):

$$RIP = O / E = 40 / 20 = 2,0$$

(La RIP también puede calcularse por división de la proporción de cánceres de pulmón entre los impresores por la proporción de cánceres de pulmón en la región (es decir, $RIP = 0,40 / 0,20 = 2,0$)).

Así, el número de cánceres de pulmón observados entre los impresores fue el doble del número esperado si los impresores presentaran la misma proporción de cánceres de pulmón (respecto a todos los cánceres) que la población general de la región entera.

alidad) estandarizadas, a excepción de que en el cálculo de los números esperados se toman como el estándar una serie de proporciones específicas por edad en lugar de unas tasas específicas por edad (véase Apartado 4.3.3).

Puede demostrarse que las *odds ratios* pueden ser más adecuadas que las razones de incidencia (o mortalidad) proporcional en situaciones en las que no se dispone de denominadores adecuados (Miettinen y Wang, 1981). Las *odds ratios* pueden calcularse para estimar el riesgo de un cáncer X particular ('casos') respecto a otros tipos de cáncer ('controles') en un grupo A de la población ('expuestos') comparado con otro grupo B de la población ('no expuestos').

En el Ejemplo 11.15, se examinaron los datos de mortalidad en una serie de análisis caso-control. En cada uno de estos análisis, todos los cánceres excepto el concreto en estudio se tomaron como 'controles' y las personas nacidas en Inglaterra y Gales como los 'no expuestos'.

En el Ejemplo 11.7 se utilizó un método similar para examinar las diferencias de educación en la mortalidad por cáncer en São Paulo (Brasil). Aunque los datos sobre el nivel educativo estaban recopilados en el censo, era probable que la calidad de estos datos fuera diferente de la de los certificados de defunción ya que la información del censo la habían proporcionado los propios individuos, mientras que en los certificados de defunción la facilitaron los parientes del fallecido. Para superar la falta de comparabilidad entre los datos del numerador y del denominador, se calcularon las *odds ratios* en lugar de las tasas.

La discusión sobre los méritos relativos de las *odds ratios* frente a las razones de incidencia o mortalidad proporcionales está fuera del objetivo de este capítulo. Basta decir que puede demostrarse que la *odds ratio* equivale a la razón de tasas, la medida que se habría calculado si se hubiera dispuesto de denominadores adecuados, siempre que la tasa de incidencia total de 'otros cánceres' sea similar en los dos grupos de población, es decir, siempre que la tasa de incidencia total de 'otros cánceres' en el análisis no esté relacionada con la 'exposición' (Miettinen y Wang, 1981). No existe una relación similar entre las razones proporcionales y las razones de tasas. A pesar de esta ventaja de las *odds ratios*, cabe destacar que el cálculo de las *odds ratios* no resuelve un problema importante de las razones proporcionales —el 'efecto de préstamo' que se produce en poblaciones con una alta incidencia de un cáncer especialmente frecuente, que da lugar inevitablemente a una menor incidencia proporcional de los otros cánceres. En otras palabras, la proporción (o *odds*) de una causa determinada puede ser alta bien porque la incidencia de esta causa es alta, o bien porque la incidencia de otras causas importantes es baja. Estas dos situaciones sólo pueden distinguirse cuando se dispone de datos adecuados del denominador (véase Apartado 4.3.5).

11.1.7 Interpretación

Los sistemas de vigilancia rutinarios suelen incluir grandes poblaciones, a veces de millones de personas seguidas durante largos periodos de tiempo. Por tanto, una de las principales ventajas de los estudios basados en da-

Ejemplo 11.15. En Inglaterra y Gales, desde 1969, en los certificados de defunción se ha incluido información sobre el país de nacimiento de los fallecidos y la Oficina de Censos y Encuestas de Población [Office of Population Censuses and Surveys (OPCS)] ha codificado esta información y la ha incluido en los archivos nacionales de mortalidad. La información sobre la raza no se registra en el certificado de defunción pero, para las muertes desde 1973 a 1985, la OPCS codificó el origen étnico para las personas nacidas en la India. La codificación se basó en el análisis del nombre, lugar exacto de nacimiento y otros datos del certificado de defunción y separó, para los nacidos en la India, los de origen étnico hindú, los de origen étnico británico y otros. En ese momento en el censo no se recopilaban datos sobre el grupo étnico y por tanto no se disponía de denominadores adecuados. De este modo, se calcularon las odds ratios ajustadas por edad para estimar los riesgos relativos de mortalidad para los diferentes tipos de cáncer en los inmigrantes en comparación con los nacidos y residentes en Inglaterra y Gales (Swerdlow et al., 1995). La Tabla 11.5 muestra los resultados de este análisis para algunos tipos de cáncer.

Tabla 11.5. Odds ratios de mortalidad ajustadas por edad para algunos tipos de cáncer seleccionados en varones emigrantes de la India a Inglaterra y Gales, por grupo étnico, en comparación con los varones nativos de Inglaterra y Gales, 1973-85^a.

Tipo de cáncer (Código CIE-9)	Nativo de Inglaterra y Gales ^b		Etnia británica, nacido en la India		Etnia hindú, nacido en la India	
	Nº	Odds ratio	Nº	Odds ratio (IC 95%) ^c	Nº	Odds ratio (IC 95%) ^c
Oral (141, 143-145)	4.564	1,0	24	1,7 (1,1-2,5)	30	2,2 (1,5-3,1)
Faringe (excepto nasofaringe) (146, 148,149)	3.440	1,0	22	2,1 (1,4-3,1)	55	5,5 (4,2-7,2)
Hígado (155)	6.177	1,0	33	1,7 (1,2-2,5)	88	5,0 (4,0-6,2)
Pulmón (163)	296.012	1,0	759	0,7 (0,7-0,8)	581	0,6 (0,6-0,7)

^a Datos de Swerdlow et al. (1995).
^b Tomado como el grupo 'no expuesto'.
^c IC = Intervalo de confianza

Así por ejemplo, para el cáncer de pulmón en los inmigrantes de etnia hindú, se construyó la Tabla 2x2 siguiente:

	Etnia hindú, nacido en la India	Nacido en Inglaterra y Gales
Cáncer de pulmón ('casos')	a (581)	b (296.012)
Otros cánceres ('controles')	c (1.737)	d (466.756)
OR = (a/b)/(c/d) = (581/296.012)/(1.737/466.756)=0,5		
(Para ajustar la edad se utilizaron los métodos descritos en el Capítulo 14.)		

tos rutinarios es que permiten el estudio de un gran número de personas a un coste muy bajo.

No obstante, una de sus limitaciones es que el número de variables recopiladas por estos sistemas suele ser limitado y con frecuencia tienden a ser

sólo medidas aproximadas de las exposiciones biológicamente más importantes. Por ejemplo, el país de nacimiento puede interpretarse como marcador de factores ambientales (p.e. dieta, variables reproductoras) para los cuales no se dispone de datos en los sistemas de datos rutinarios.

Calidad de los datos

En la interpretación de los resultados de estos estudios deben considerarse los artefactos potenciales de los datos. La observación de diferencias en la incidencia del cáncer registrada entre poblaciones no refleja necesariamente unas variaciones subyacentes reales en los riesgos de cáncer. Las diferencias en la incidencia registrada pueden aparecer debido a diferencias en el acceso al servicio sanitario (incluyendo el cribaje), al diagnóstico y a las prácticas de registro. Finalmente, las diferencias en la incidencia pueden ser debidas a variaciones en la exactitud del recuento de la población. Algunos de estos aspectos se comentaron brevemente en el Apartado 2.9 y el Anexo A2.2.

Las tendencias crecientes en la incidencia de cáncer de próstata en muchos países occidentales han sido difíciles de evaluar debido a que no se sabe en qué proporción han estado afectadas por cambios en las prácticas de diagnóstico y registro. Se han observado fuertes paralelismos en los EE.UU. entre las tendencias en el tiempo de la incidencia del cáncer de próstata en estadio local y las tendencias en el tiempo del uso de la resección transuretral de la próstata (Potosky *et al.*, 1990; Severson, 1990), un procedimiento quirúrgico efectuado para aliviar síntomas frecuentes de hiperplasia prostática benigna. El uso creciente de este procedimiento médico parece haber dado lugar a una mayor detección de tumores clínicamente inaparentes que no se habrían diagnosticado de otro modo. Las variaciones entre países en el uso de esta técnica también parecen influir en algunas de las diferencias en las tasas de incidencia en Japón y los EE.UU. (Shimizu *et al.*, 1991).

En muchos países no están disponibles los datos de incidencia del cáncer o, si lo están, el sistema de registro no cubre toda la población o los datos obtenidos son incompletos. Los datos de mortalidad están disponibles para muchos más países y durante periodos de tiempo más prolongados. No obstante, los datos de mortalidad presentan el inconveniente de que no reflejan directamente la incidencia, especialmente para cánceres con un buen pronóstico. Esto es debido a que la mortalidad depende de la incidencia y de la letalidad de los casos. Las tasas de mortalidad son prácticamente paralelas a las tasas de incidencia sólo si la enfermedad es fatal y si la muerte se produce poco después del diagnóstico. Por ejemplo, las tasas de mortalidad son una buena indicación de la magnitud de las tasas de incidencia para el cáncer de pulmón porque este tumor presenta una fatalidad alta y rápida pero no para el cáncer de testículo que presenta una supervivencia relativamente alta. La exactitud de los datos de mortalidad proporcionados en los certificados de defunción también está influenciada por las diferencias geográficas y las tendencias en las prácticas de diagnóstico y

certificación y por cambios en la frecuencia con la que se efectúan los exámenes post-mortem.

Las estadísticas rutinarias de mortalidad en muchos países se basan en una única causa subyacente de muerte. Las normas para la determinación de la causa subyacente de muerte son principalmente las de la *Clasificación Internacional de Enfermedades* (véase Anexo A.2.2), pero en algunas ocasiones las oficinas nacionales de estadísticas vitales pueden imponer sus propias normas. Por ejemplo, la Oficina de Censos y Encuestas de Población en Inglaterra y Gales cambió una de estas normas (norma 3) en 1983, de modo que es más probable que el cáncer se codifique a partir de este momento como la causa de muerte incluso aunque en el certificado de defunción sólo se registre como una causa que contribuye (Ashley y Devis, 1992). Este cambio en la codificación afectó a las tendencias seculares de la mortalidad por cáncer. Se ha estimado que cerca de un 50% del aumento de la mortalidad por cáncer de próstata en las edades de 75-84 en Inglaterra y Gales desde 1970 a 1990 podrían explicarse por cambios en las prácticas de codificación (Grulich *et al.*, 1995).

Otros muchos factores, algunos de ellos menos evidentes, pueden afectar a la incidencia registrada y a los datos de mortalidad. Las tasas de mortalidad por neoplasias cerebrales malignas disminuyeron mucho en Inglaterra y Gales durante los años 1981-1982, pero a partir del examen de los datos de neoplasias cerebrales de naturaleza no especificada, está claro que esta disminución era en gran parte, o totalmente, un artefacto. La mayoría de los casos de naturaleza no especificada se codifican como tumores cerebrales después de haber solicitado a los autores de los certificados una aclaración sobre la causa de muerte. El artefacto mostrado en la Figura 11.9 se produjo cuando se redujo el número de solicitudes de aclaración a los certificadores durante la huelga de registradores en 1981-82.

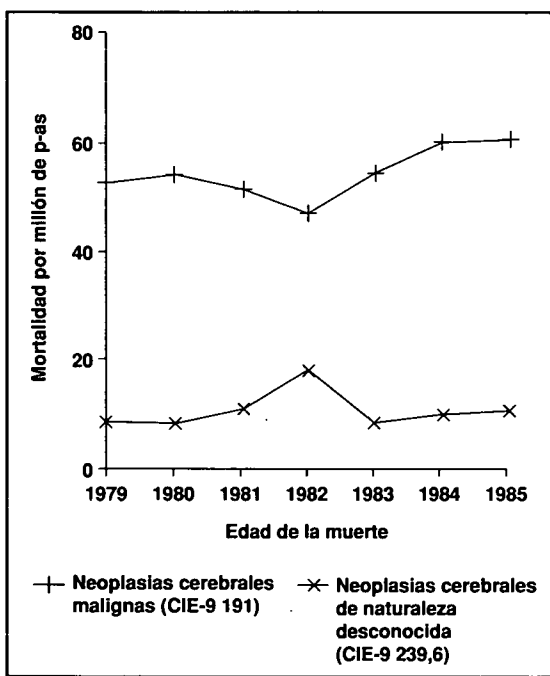


Figura 11.9. Mortalidad por neoplasias cerebrales, malignas y de naturaleza no especificada, en hombres, Inglaterra y Gales, 1979-85. Tasas de mortalidad estandarizadas por edad a la población de Inglaterra y Gales en 1981 (reproducido con la autorización de Churchill Livingstone, de Swerdlow (1989)).

La mayoría de los casos de naturaleza no especificada se codifican como tumores cerebrales después de haber solicitado a los autores de los certificados una aclaración sobre la causa de muerte. El artefacto mostrado en la Figura 11.9 se produjo cuando se redujo el número de solicitudes de aclaración a los certificadores durante la huelga de registradores en 1981-82.

Comparabilidad de los datos del numerador y del denominador

Los cálculos de tasas requieren estimaciones de personas-años a riesgo, según las variables en estudio. Éstas normalmente se obtienen de los datos del censo. Es esencial que las variables en el censo estén definidas, clasificadas y codificadas de un modo similar a las variables en las bases de datos de los registros del cáncer y de mortalidad. No obstante, incluso

cuando se utilizan los mismos criterios, los individuos todavía pueden clasificarse de manera diferente en las dos bases de datos si difiere el modo como se obtuvo la información. Por ejemplo, la información en el censo la proporcionan los propios individuos mientras que en los certificados de defunción la proporcionan los informadores (normalmente parientes del fallecido). Puede que los informadores no sean capaces de proporcionar una información exacta sobre muchos aspectos tales como la ocupación

del fallecido, o pueden indicar intencionadamente una ocupación más prestigiosa (OPCS, 1978).

Sesgo de selección

Las personas que migran generalmente no son un grupo representativo de sus poblaciones de origen. Por ejemplo, los emigrantes tienden a estar más sanos que los que permanecen en el país de origen. Este 'efecto del emigrante sano' afectará a las comparaciones con los riesgos de la población origen. Un sesgo similar puede producirse en el análisis por ocupación (el 'efecto del trabajador sano') y esto afectará a las comparaciones con la población general (véanse los Apartados 8.2.2 y 13.11).

Otra fuente de sesgo más difícil de detectar surge de cambios en la 'exposición' relacionados ó a consecuencia de la propia enfermedad. Por ejemplo, cuando los emigrantes vuelven a su país de origen poco antes de la muerte o cuando las personas cambian de trabajo como consecuencia del diagnóstico de una enfermedad particular. En estas situaciones, los riesgos en el país huésped o en el trabajo realizado antes del diagnóstico estarán infraestimados.

Confusión

Una de las principales limitaciones de los estudios basados en datos rutinarios es que generalmente no se dispone de los datos sobre factores de confusión importantes (a excepción de la edad y el sexo). Por ejemplo, muchos de los riesgos elevados de cáncer de pulmón encontrados en muchos grupos ocupacionales no son debidos a exposiciones ocupacionales sino a una alta prevalencia de tabaquismo. Lamentablemente, los datos sobre los hábitos de fumar raramente se recopilan en los sistemas rutinarios de vigilancia.

Consideraciones finales

En resumen, el estudio de las variaciones en la incidencia y mortalidad del cáncer por lugar de residencia y nacimiento, grupo étnico, nivel socio-económico y en el tiempo, es un ejercicio válido y útil, siempre que el investigador posea un amplio conocimiento de cómo se recopilan y procesan los datos, de modo que en la interpretación de los resultados se consideren todas las fuentes posibles de artefactos, sesgos y confusión de datos.

11.2 Datos agregados (estudios ecológicos)

Los estudios basados en datos rutinarios considerados hasta ahora comparten la característica de que las observaciones realizadas pertenecen a sujetos individuales. No obstante, puede investigarse a nivel de grupos en lugar de a nivel individual. Los estudios que implican investigar la frecuencia de la enfermedad (o de cualquier otro resultado de interés) respecto al nivel de exposición en varios grupos de individuos (o en el mismo grupo en diferentes periodos de tiempo) se denominan estudios ecológicos. En este tipo de estudio, no es posible relacionar la exposición de un individuo en particular con su estado de salud. Por tanto, la unidad de observación y análisis es el grupo en lugar del individuo. Los grupos pueden definirse de

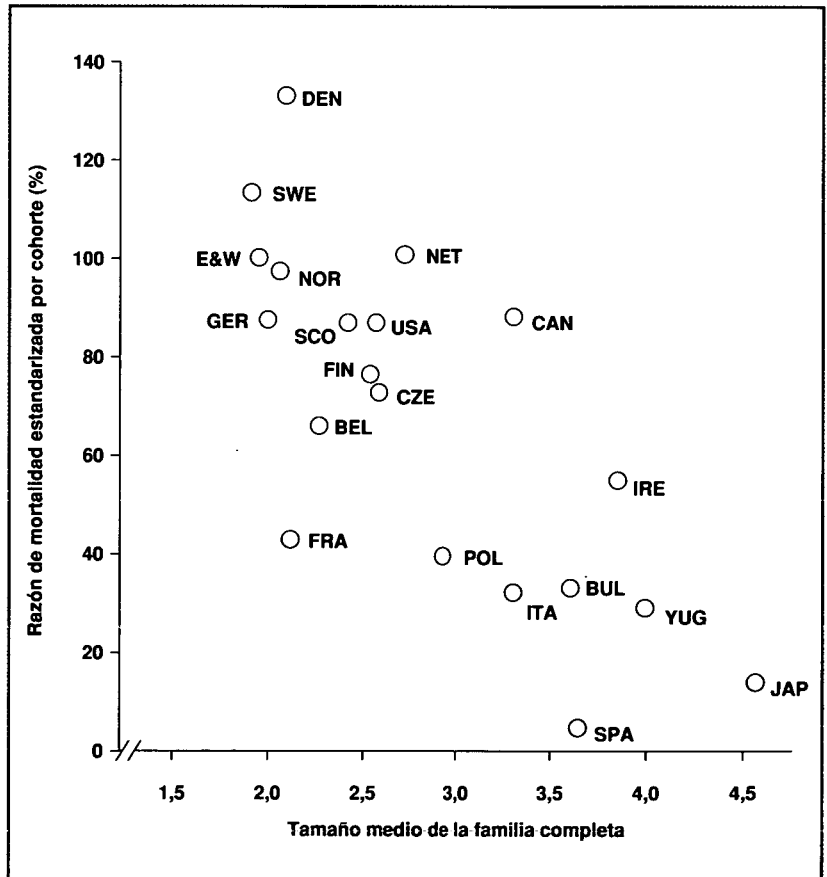
muchas formas según el lugar de residencia, lugar de nacimiento, nivel socioeconómico, ocupación, etc.

Los estudios ecológicos se utilizan con frecuencia como un primer paso en la investigación de una posible relación exposición-enfermedad debido a que pueden realizarse rápida y económicamente utilizando información ya disponible. Los datos de exposición pueden obtenerse a partir de organizaciones gubernamentales y privadas que recopilan rutinariamente datos sobre variables demográficas, ambientales y del estilo de vida. Las tasas de enfermedad pueden obtenerse de las oficinas de estadísticas vitales, programas de vigilancia o registros de enfermedad (p.e. registros del cáncer).

Ejemplo 11.16. Los datos recopilados rutinariamente sobre la mortalidad por cáncer de ovario en veinte países se examinaron respecto al tamaño medio de la familia, estimado a partir de varios estudios demográficos. Había una clara relación inversa entre la mortalidad por este tumor y el tamaño medio de la familia completa, es decir, el número total de niños por mujer al final de su vida reproductiva (Figura 11.10) (Beral et al., 1978). Esta observación indicaba que el embarazo protegía del cáncer de ovario, una hipótesis que se ha confirmado en muchos estudios caso-control y de cohorte.

Figura 11.10.

Diagrama de dispersión de la relación entre las tasas de mortalidad estandarizadas por edad por cáncer de ovario y el tamaño medio de la familia completa para mujeres de 20 países nacidas alrededor de 1901. Las tasas de mortalidad específicas de edad en Inglaterra y Gales se utilizaron como estándar, es decir, la razón de mortalidad para Inglaterra y Gales = 100 (reproducido con la autorización de Beral et al., 1978). © The Lancet Ltd., 1978). BEL = Bélgica; BUL = Bulgaria; CAN = Canadá; CZE = antigua Checoslovaquia; DEN = Dinamarca; EYW = Inglaterra y Gales; FRA = Francia; FIN = Finlandia; GER = Antigua Alemania Oriental; IRE = Irlanda; ITA = Italia; JAP = Japón; NET = Holanda; NOR = Noruega; POL = Polonia; SCO = Escocia; SPA = España; SWE = Suecia; USA = Estados Unidos de América; YUG = antigua Yugoslavia.



En el Ejemplo 11.16, no es posible relacionar la mortalidad por cáncer de ovario de una mujer individual con el tamaño de su familia debido a que la única información disponible incluía el tamaño medio de sus familias y una estimación de la mortalidad media por cáncer de ovario para cada país incluido en el análisis. Por tanto, la unidad del estudio fue el país en lugar del individuo.

Pueden realizarse comparaciones similares entre el cambio en el tiempo de la exposición media y el cambio en la tasa de enfermedad para una población individual (como en el Ejemplo 11.16).

Ejemplo 11.17. En el estudio sobre la mortalidad por cáncer de ovario y el tamaño de la familia descrito en el Ejemplo 11.16, en el Reino Unido y los EE.UU., se demostró que la mortalidad por cáncer de ovario observada en cohortes de nacimiento de mujeres sucesivas, estaba estrechamente relacionada con el tamaño de su familia (Figura 11.11).

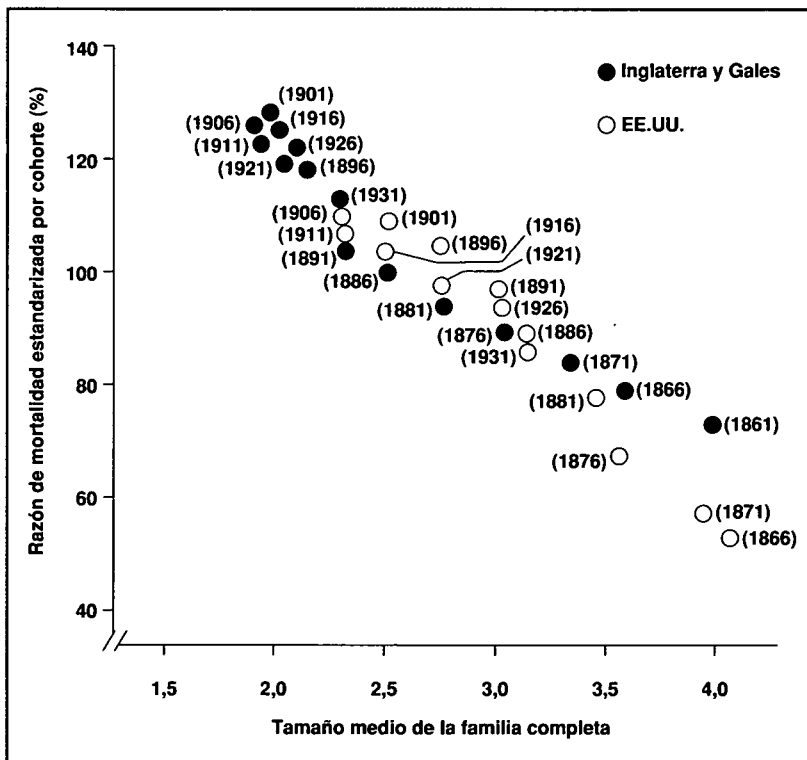


Figura 11.11.

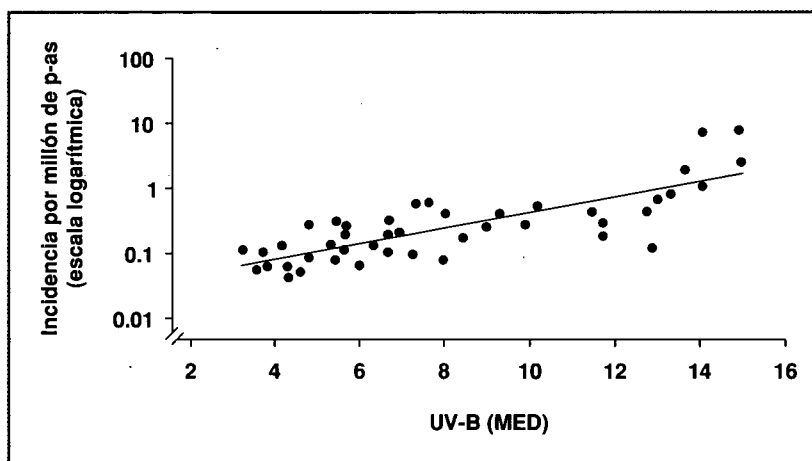
Diagrama de dispersión de la relación entre las tasas de mortalidad estandarizadas por edad por cáncer de ovario y el tamaño medio de la familia completa para diferentes generaciones de mujeres nacidas en Inglaterra y Gales y los EE.UU. La estructura de edad promedio de la población combinada de Inglaterra, Gales y los EE.UU. entre 1931 – 1973 se tomó como estándar. Por tanto, una razón de 140 significa que la mortalidad por cáncer de ovario de la cohorte es del 40% mayor que el promedio para mujeres en Inglaterra, Gales y los EE.UU. (el año medio de nacimiento de cada generación se indica entre paréntesis) (reproducido con la autorización de Beral *et al.*, 1978. © The Lancet Ltd., 1978).

Los estudios ecológicos pueden ser el diseño más adecuado para estudiar exposiciones que sean más fáciles de definir y medir en una población que a nivel individual. Este es el caso de muchas exposiciones ambientales tales como la contaminación del aire, calidad del agua y radiación ultravioleta (Ejemplo 11.18).

Ejemplo 11.18. Se obtuvieron datos recopilados rutinariamente por numerosos registros del cáncer de base poblacional y mediciones publicadas de intensidad de luz ultravioleta solar ambiental para evaluar si la distribución geográfica del carcinoma de células escamosas del ojo estaba relacionado con la luz ultravioleta solar. El análisis se basó en datos de 47 poblaciones: 3 en África, 9 en Australia, 20 en Europa, 12 en Norteamérica, 2 en Sudamérica, 1 en el Oriente Medio. El periodo de estudio cubierto por cada registro varió pero la mayoría incluía los años 80. Los resultados de este estudio (Figura 11.12) se correspondían con la hipótesis de que la exposición a la luz ultravioleta solar aumenta el riesgo de carcinoma de células escamosas del ojo (Newton, et al., 1996).

Figura 11.12.

Tasas de incidencia ajustadas por edad y sexo de carcinoma de células escamosas del ojo en relación a las determinaciones de radiación ultravioleta B expresado como dosis de eritema mínima (MED), una unidad que refleja mejor la dosis biológicamente eficaz (reproducido con la autorización de Newton *et al.*, 1996, © por The Lancet Ltd., 1978).



Los estudios ecológicos también son útiles para monitorizar la eficacia de intervenciones sobre la población tales como campañas de educación sanitaria (p.e. campaña antitabaco), programas de inmunización o programas generalizados de cribaje.

11.2.1 Análisis

Los estudios ecológicos difieren de los estudios epidemiológicos individuales en que los individuos ‘expuestos’ y los ‘no expuestos’ en cada una de las poblaciones no están realmente identificados. Por tanto, no es posible determinar la fuerza de la asociación entre la exposición y el resultado utilizando alguna de las aproximaciones descritas en el Capítulo 5 para estudios basados en individuos.

Además, a diferencia de otros estudios epidemiológicos, el resultado de un estudio ecológico es normalmente una variable cuantitativa (p.e. tasa de mortalidad) en lugar de una categoría binaria (por ejemplo, ‘enfermo’ ó ‘no enfermo’). La variable de exposición también tiende a determinarse en una escala cuantitativa. Incluso las variables cualitativas se convierten en cuantitativas cuando se promedian para una población: el sexo es una va-

riable binaria, pero la proporción de una población que es masculina, o femenina, es una variable cuantitativa.

Una medida estadística denominada *coeficiente de correlación*, indicada por r , se ha utilizado ampliamente en estudios ecológicos para medir la fuerza de la asociación entre la exposición y la enfermedad. Otro método más adecuado es ajustar una recta de *regresión* que predice la incidencia o mortalidad como una función del nivel de exposición. Antes de comentar las ventajas y limitaciones de cada uno de estos procedimientos, estos dos métodos estadísticos se introducen aquí brevemente.

Introducción a los conceptos de regresión y correlación

Al investigar la relación entre dos variables cuantitativas el primer paso debe ser siempre representar gráficamente los datos (véase Apartado 3.3.3). La representación necesaria se denomina *diagrama de dispersión* en el que el eje vertical (y) se refiere a la variable resultado y el eje horizontal (x) a la variable exposición (como en las Figuras 11.10-1.12). En este tipo de gráfico, cada grupo considerado se representa por un punto.

Consideremos el diagrama de dispersión de la Figura 11.13(a). Los puntos en el diagrama de dispersión muestran una tendencia clara, hacia arriba y a la derecha; se dice que existe una relación *positiva* entre las dos variables. Valores altos (bajos) de una variable están asociados con valores altos (bajos) de la otra. Podemos desear resumir la relación entre las dos variables, y por tanto poder predecir el valor de una variable cuando sólo conocemos la otra variable. Una posibilidad de resumir esta relación es simplemente trazar una línea recta por los puntos de dispersión. Toda línea recta trazada en una gráfica puede representarse por una ecuación. En el ejemplo anterior esta relación podría resumirse como $y = x$, porque la recta se encuentra exactamente a 45 grados. Si el valor de y fuera siempre 0,9 del valor de x (es decir, si la recta se encontrara a un ángulo ligeramente menor), la relación podría expresarse como $y = 0,9x$.

Más generalmente, la ecuación para una línea recta se expresa como:

$$y = a + bx$$

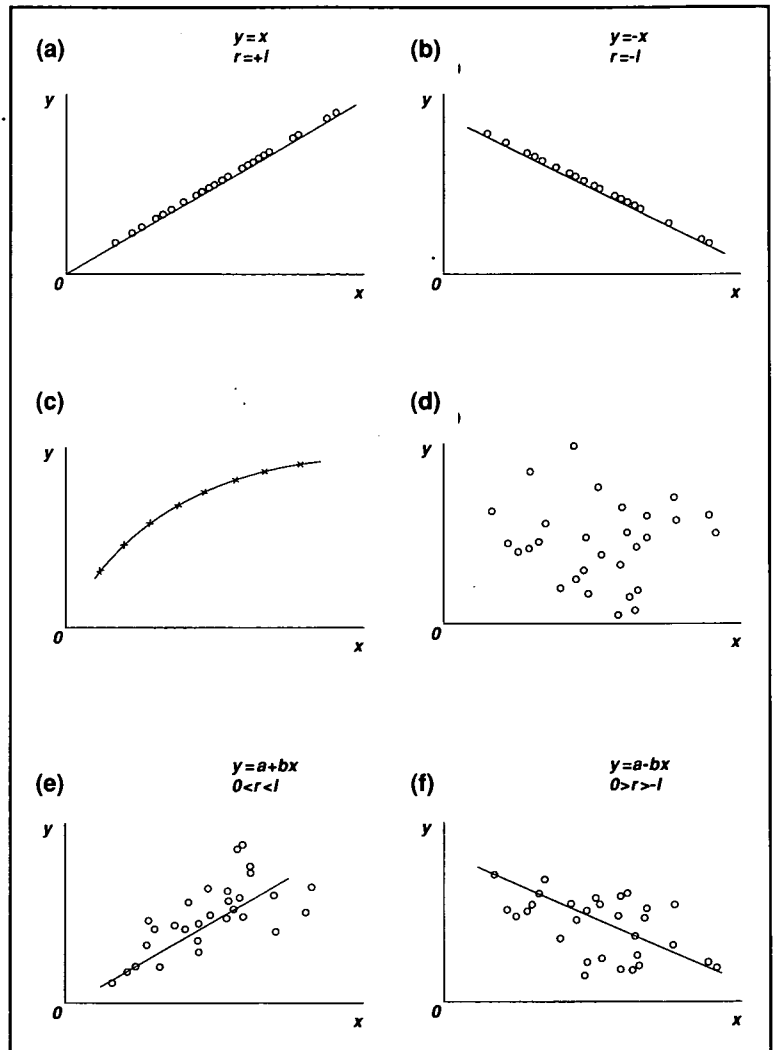


Figura 11.13. Diagramas de dispersión, ecuaciones de regresión y coeficientes de correlación.

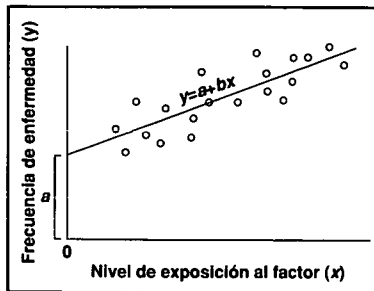


Figura 11.14.
Diagrama de dispersión y recta de regresión

donde y se refiere a los valores de la variable resultado y x a los valores de la variable exposición, a es la intersección de la recta con el eje y , y b es la pendiente de la recta, el aumento (o disminución) de y por cada unidad de aumento de x (Figura 11.14). En el caso especial en que la recta pasa por el origen (O) de los dos ejes, como en la Figura 11.13(a) y la recta se encuentra exactamente a 45 grados, la ecuación se reduce a $y = x$, ya que $a = 0$ y $b = 1$.

En la Figura 11.13(b), también existe una asociación perfecta entre x e y , pero la tendencia es disminuir hacia la derecha; se dice que existe una relación *inversa*, o *negativa*, entre las dos variables. Esta relación también puede expresarse por la ecuación $y = a + bx$, excepto que en este caso el coeficiente b tomaría un valor negativo, indicando que cuando una variable aumentaba, la otra disminuía.

En las Figuras 11.13(a) y (b), los puntos en el diagrama de dispersión se encuentran en una línea recta y se dice que la relación es *lineal*. En la Figura 11.13(c), todavía existe una clara relación entre las dos variables pero la relación presenta una forma *no lineal* y la ecuación es más compleja. Por último, en el Ejemplo 11.13(d), no existe una relación evidente entre x e y . Valores altos de x están asociados con valores altos y bajos de y . Los puntos en el diagrama de dispersión no muestran una tendencia especial y en ausencia de una relación entre las dos variables, no es apropiado utilizar una ecuación.

Las relaciones ilustradas en las Figuras 11.13(a), (b) y (e) son perfectas debido a que todos los puntos en el diagrama de dispersión se encuentran en una línea. No obstante, en la mayoría de las situaciones reales, los puntos estarán dispersos a su alrededor (como en las Figuras 11.13(e) y (f)). En estas circunstancias, utilizamos una línea recta que proporcione la 'mejor' predicción de y para cualquier valor de x . Podríamos trazar una línea 'a ojo' pero este método sería subjetivo y sería poco probable que proporcionara la mejor línea. Una alternativa es utilizar un método estadístico especial, denominado *análisis de regresión*, para encontrar la mejor recta que se 'ajuste' a los datos. La ecuación de la recta obtenida por este método se denomina *ecuación de regresión* (en Armitage y Berry (1994) se presenta un ejemplo de los cálculos).

Este método estadístico se aplicó a los datos de mortalidad por cáncer de ovario mostrados en la Figura 11.11 (utilizando los datos de los dos países) y se obtuvo la siguiente ecuación de regresión (Beral *et al.*, 1978):

$$y = 182 - 30x$$

Esta ecuación significa que para un valor dado de x (es decir, tamaño medio de la familia), puede calcularse un valor asociado para y (es decir, mortalidad por cáncer de ovario). Por tanto, la razón de mortalidad estandarizada por edad de cáncer de ovario en una cohorte de nacimiento particular de mujeres puede predecirse a partir de esta ecuación. Por ejemplo, la razón de mortalidad en una cohorte con un tamaño medio de la familia de 3 niños puede calcularse por sustitución de $x = 3$ en la ecuación de regresión

$$y = 182 - 30 \times 3 = 92$$

que es el valor que habríamos estimado al examinar la gráfica. Este valor predicho puede interpretarse como el valor *medio* de y asociado a un valor dado de x .

Si todos los valores en el diagrama de dispersión se encontraran en la recta de regresión, los valores de y esperados y observados serían idénticos. La ecuación de regresión describiría exactamente la relación entre el tamaño medio de la familia y la mortalidad por cáncer de ovario. En otras palabras, la variación en la mortalidad por cáncer de ovario se explicaría completamente por las variaciones en el tamaño de la familia. En la práctica no es así. La mortalidad por cáncer de ovario puede variar independientemente de las variaciones del tamaño de la familia, de modo que dos poblaciones con el mismo tamaño medio de la familia pueden presentar riesgos de mortalidad por cáncer de ovario diferentes. Por tanto, la ecuación de regresión sólo puede medir la relación *media* entre las dos variables.

Si los puntos en el diagrama de dispersión se encuentran cerca de la recta de regresión, esto indica que los valores observados de y no difieren considerablemente de los valores predichos representados por la recta de regresión. Por tanto, la mayor parte de la variación en y puede explicarse por la variación en x . Por otro lado, si existe una dispersión importante de los puntos alrededor de la recta de regresión, una parte considerable de la variación en y no se explica por la variación en x .

Para cuantificar la magnitud de la dispersión alrededor de la recta de regresión, podemos calcular una medida denominada coeficiente de correlación, r . El valor de este coeficiente siempre se encuentra entre -1 y $+1$:

- (a) para correlaciones *positivas* perfectas (Figura 11.13(a)), $r = +1$.
- (b) para correlaciones *negativas* perfectas (Figura 11.13(b)), $r = -1$.
- (c) si existe alguna dispersión alrededor de la recta de regresión (Figuras 11.13(e), (f)), r se encuentra entre 0 y $+1$ (o entre 0 y -1). Cuanto menor sea la dispersión más próximo es r a 1 (o a -1).
- (d) si no existe una relación *lineal* entre y y x , r es próximo a 0 . Esto implica o bien que no existe una relación entre las dos variables (Figura 11.13d) o bien que la relación es *no lineal* (Figura 11.13c).

Utilización de la correlación y la regresión en el análisis de estudios ecológicos

El coeficiente de correlación se utiliza con bastante frecuencia en el análisis de estudios ecológicos como medida de la fuerza de la asociación entre exposición y enfermedad. No obstante, no es la estrategia más adecuada. En primer lugar, su magnitud depende del rango de valores de la variable de exposición; si éste es amplio, la correlación será mayor que si es estre-

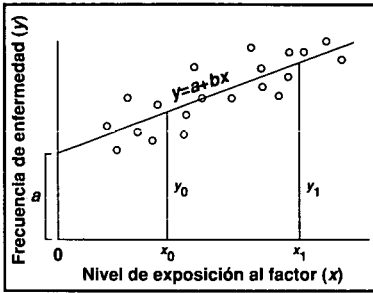


Figura 11.15. Representación de la asociación entre frecuencia de la enfermedad (y) y nivel de exposición (x) al factor estudiado en diferentes poblaciones.

cho. En segundo lugar, el valor del coeficiente de correlación no puede traducirse en ninguna de las medidas convencionales de riesgo relativo y , por tanto, es difícil de interpretar en términos epidemiológicos.

Un método más adecuado para cuantificar el efecto de la exposición en un estudio ecológico es ajustar una recta de regresión a los datos, la cual predice la incidencia o mortalidad en función del nivel (o prevalencia) de la exposición. A diferencia del coeficiente de correlación, este método no se ve afectado por cambios en el rango de la variable exposición. Además, puede estimarse una medida del riesgo relativo a partir de la pendiente y la intersección de la recta de regresión (Beral *et al.* 1979).

Supongamos que en n poblaciones diferentes el nivel de enfermedad (y) está relacionado linealmente con el nivel de exposición medio (x) al factor a estudiar. Esta relación puede expresarse por una ecuación de regresión ajustada a los datos de n poblaciones

$$y = a + bx$$

Podemos considerar dos de estas poblaciones con diferentes niveles medios de exposición x_0 y x_1 y diferentes frecuencias de enfermedad asociadas y_0 y y_1 , respectivamente (Figura 11.15). El riesgo relativo entre las dos poblaciones puede definirse como la frecuencia de la enfermedad en individuos con una exposición media x_1 , respecto a los individuos con una exposición media x_0 , o simplemente y_1/y_0 .

Así, si

$$y_0 = a + bx_0 \quad \text{y} \quad y_1 = a + bx_1$$

la estimación del riesgo relativo (RR) puede calcularse como

$$RR = \frac{y_1}{y_0} = \frac{a + bx_1}{a + bx_0}$$

En nuestro ejemplo de mortalidad por cáncer de ovario, la ecuación de regresión fue igual a $y = 182 - 30x$. Por tanto, el riesgo relativo de cáncer de ovario en mujeres con una media de 2 niños en comparación con aquellas con una media de 4 niños puede estimarse como:

$$RR = \frac{182 - 30 \times 2}{182 - 30 \times 4} = 2,0$$

Si la exposición se mide en términos de la *proporción* de personas en la población expuestas al factor de interés (p.e. proporción de fumadores) en lugar de hacerlo en términos del *nivel medio* de la exposición (p.e. promedio de niños) como en el ejemplo anterior, el riesgo relativo puede estimarse como:

$$RR = \frac{\text{Pendiente}}{\text{Intersección}} + 1 = \frac{b}{a} + 1$$

Esta fórmula es sólo un caso especial de la ecuación anterior obtenida sustituyendo $x_0 = 0$, $x_1 = 1$, que son el intervalo de valores que puede tomar una proporción.

Este método incluye algunas suposiciones. En primer lugar, supone que la variable y (frecuencia de enfermedad) está relacionada linealmente con la variable x (exposición). En segundo lugar, supone que la frecuencia de la enfermedad en cada grupo de población se determina completamente por el nivel de exposición. En tercer lugar, supone que la exposición se mide sin error. Como veremos más adelante, muchas veces estas suposiciones no se cumplen, lo que limita el uso e interpretación de estos métodos.

11.2.2 Interpretación de los estudios ecológicos

Falacia ecológica

La observación de una relación a nivel de población entre dos variables no implica necesariamente que la misma relación se mantenga a nivel individual. Esto se conoce como la falacia ecológica. Por ejemplo, el análisis anterior sobre el tamaño de la familia y el cáncer de ovario indicaba que el embarazo podría conferir protección frente al cáncer de ovario. No obstante, no sabemos si las mujeres que murieron de cáncer de ovario en cada grupo de población eran realmente aquellas con menos hijos o sin hijos.

En un estudio ecológico, los niveles de exposición representan los niveles medios para cada grupo de población. Debido a que el interés real reside en el riesgo individual determinado por la exposición individual, los estudios ecológicos sólo están justificados en la situación ideal en la cual cada individuo dentro de un grupo presenta el mismo nivel de exposición. Esto es así en muchas exposiciones ambientales tales como la contaminación del aire, la radiación natural y la calidad del agua.

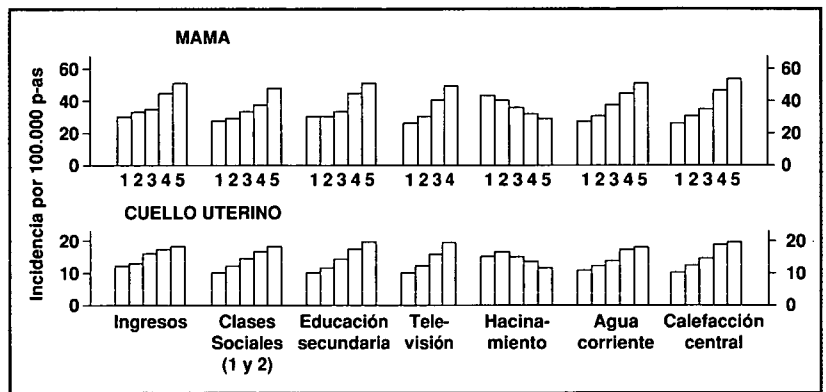
No obstante, muchas veces la exposición es heterogénea dentro de un grupo porque algunos individuos no están expuestos y es probable que los expuestos lo estén a diferentes niveles. En estas situaciones, la relación exposición-respuesta observada a nivel del grupo no refleja necesariamente la relación exposición-respuesta a nivel individual. Por tanto, la observación de una relación lineal entre la exposición media y la frecuencia de la enfermedad en un estudio ecológico no implica que esta relación lineal este presente a nivel individual.

En el análisis ecológico presentado en el Ejemplo 11.19, el riesgo de cáncer de mama y de cuello uterino aumentó al aumentar el nivel socioeconómico. Este gradiente positivo es lo que se esperaría para el cáncer de mama pero no para el cáncer de cuello uterino (véanse Ejemplos 11.7 y 11.9).

Cuando se realizaron análisis a nivel individual, las tendencias socioeconómicas y de educación en el riesgo fueron diferentes para los cánceres de mama y de cuello uterino (Ejemplo 11.20). Las mujeres con mejores traba-

Ejemplo 11.19. Se agruparon aproximadamente 500 municipios finlandeses en diferentes categorías (cinco o cuatro) según varios indicadores socioeconómicos tales como ingresos mensuales medios por habitante en 1968, porcentaje de habitantes que pertenecen a las dos clases sociales más altas (clases sociales 1 y 2), porcentaje de personas con educación secundaria en 1970, número de licencias de televisión por 1.000 habitantes en 1970, porcentaje de personas que viven en casas con más de dos habitantes por habitación en 1950, y porcentaje de viviendas con agua corriente, calefacción central y electricidad en 1950. Para cada uno de estos grupos de municipios se calculó la incidencia de cánceres de mama y de cuello uterino (Hakama et al., 1982). Los resultados se presentan en la Figura 11.16.

Figura 11.16. Tasas de incidencia estandarizadas por edad de los cánceres de mama y de cuello uterino en grupos de municipios definidos según diferentes estimaciones del nivel socioeconómico, Finlandia, 1955-74. Se tomó como estándar la población media de Finlandia en 1955-74 (reproducido con la autorización de Hakama et al., 1982).



jos y nivel de educación presentaban un riesgo alto de cáncer de mama y riesgo bajo de cáncer de cuello uterino (Tabla 11.6).

La diferencia en los resultados de cáncer de cuello uterino entre los análisis individual y ecológico puede ser debida al hecho de que, por ejemplo, las mujeres con riesgo mayor de cáncer de cuello uterino (p.e. prostitutas) tendían a estar concentradas en municipios urbanos, que también son los que presentan una mayor proporción de residentes de las clases sociales más altas. Esto podría explicar potencialmente la relación positiva observada en el análisis ecológico entre la proporción de personas de las clases sociales más altas y el riesgo de cáncer de cuello uterino.

Confusión

Una segunda limitación importante de los estudios ecológicos es la falta de capacidad para controlar los efectos de factores potenciales de confusión. Por ejemplo, en un estudio de la ingesta diaria de grasa por cápita estimada en relación a la incidencia de cáncer de mama en mujeres de 35-64 años en 24 países, se observó una relación positiva entre estas dos variables (Armstrong y Mann, 1985). Este análisis indicaba una posible asociación entre la ingesta de grasa y el riesgo de desarrollar cáncer de mama. No obstante, el consumo elevado de grasa puede ser un mero marcador para otros

Ejemplo 11.20. En el estudio descrito en el Ejemplo 11.19, también se efectuaron análisis a nivel individual cruzando cada uno de los registros de cáncer de mama y de cuello uterino con los datos del censo de 1970. Para cada paciente con cáncer, se registraron los datos ocupacionales y de educación a partir de los datos del censo y los pacientes se agruparon en clases socioeconómicas y de educación basándose en los códigos del censo original (Hakama et al., 1982). Los resultados de este análisis individual se presentan en la Tabla 11.6.

	Cáncer de mama			Cáncer de cuello uterino		
	Observ. (O)	Esperad. ^b (E)	RIE (%) (100×O/E)	Observ. (O)	Esperad. ^b (E)	RIE (%) (100×O/E)
<i>Nivel socioeconómico^c</i>						
Empleados	172	148,9	116	24	35,0	69
Granjeros	562	705,1	80	108	165,9	65
Otros empleados autónomos	127	128,2	99	26	30,2	86
Gestores	343	229,3	150	24	53,8	45
Administrativos	1.061	842,7	126	177	197,3	90
Trabajadores cualificados	797	890,6	90	271	208,8	130
Trabajadores no cualificados	357	412,5	87	146	96,9	151
Pensionistas	1.169	1.228,2	95	313	299,3	105
<i>Nivel educativo</i>						
Primario	3.011	3.356,1	90	897	797,1	113
Secundario	975	847,5	115	150	200,0	75
Bachiller	255	186,7	137	26	44,0	59
Universidad	356	206,8	172	17	48,9	35
^a Datos de Hakama et al. (1982)						
^b Se tomó como estándar las tasas de cáncer de mama y cuello uterino específicas para la edad de la población femenina total de Finlandia de 30-69 años de edad..						
^c En Finlandia, la situación socioeconómica se basa en la ocupación. Para mujeres económicamente inactivas (p. e. amas de casa) se definió como la del cabeza de familia (normalmente el marido).						

Tabla 11.6.

Razones de incidencia estandarizadas (RIE) de cáncer de mama y de cuello uterino por nivel socioeconómico y educación en mujeres de 30-69 años. Finlandia, 1971-75^a.

factores relacionados con un alto riesgo de cáncer de mama, tales como un nivel socioeconómico más alto, menor fertilidad y edad avanzada al tener el primer hijo. Normalmente los estudios ecológicos no disponen de los datos sobre los factores de confusión conocidos o sospechados, pero, incluso aunque se conocieran, sería difícil utilizarlos para ajustar los resultados a nivel poblacional.

Cabe destacar que factores de riesgo que sean independientes de la exposición a nivel individual pueden correlacionarse con ella y por tanto convertirse en factores de confusión, cuando se agregan a nivel poblacio-

nal. Por ejemplo, en un estudio de la relación entre la proporción de trabajadores de la madera y el cáncer de pulmón en zonas definidas geográficamente, el hábito de fumar inducirá confusión si el consumo de cigarrillos cambia con la proporción de trabajadores de la madera en cada zona, incluso aunque los dos factores en cuestión sean independientes a nivel individual. Por el contrario, una variable de confusión a nivel individual puede no ser una variable de confusión a nivel ecológico. Por ejemplo, aunque el riesgo de cáncer de pulmón es mucho mayor para hombres que mujeres, el sexo, como variable ecológica (p.e. porcentaje de hombres), no estará asociado con la tasa de enfermedad en grupos definidos geográficamente si la proporción de hombres es similar en todos los grupos.

Errores de medida

La exposición se estima con frecuencia a partir de datos obtenidos por otras razones, los cuales generalmente sólo proporcionan una medida indirecta de los posibles factores de riesgo. Por ejemplo, los datos sobre el fumar y el consumo de alcohol muchas veces se basan en ventas que reflejan sólo parcialmente el consumo, ya que no se tienen en cuenta las pérdidas y

la importación no registrada. Además, debido a que los datos no los recopilan los propios investigadores, puede ser difícil evaluar adecuadamente su calidad.

Como se ha comentado anteriormente en el Apartado 11.1.7, la determinación de la incidencia y mortalidad por cáncer también puede estar afectada por errores.

Periodo de latencia

La mayoría de los estudios ecológicos comparan la exposición medida en un momento en el tiempo con la enfermedad medida en otro (o el mismo) momento en el tiempo. Esto se ilustra en la Figura 11.17, donde los datos sobre la tasa de enfermedad y sobre la exposición de interés (consumo de cerveza per cápita) se refieren aproximadamente al mismo periodo (1960-62).

Idealmente, en el análisis debería incorporarse un periodo de tiempo de retraso de modo que los datos de la exposición se refieran al periodo etiológico

pertinente (p.e. 10-20 años antes del desarrollo del cáncer). No siempre se dispone de los datos sobre exposiciones en el pasado y, con bastante frecuencia, se está obligado a confiar en datos demasiado recientes. Esto constituye un problema importante cuando es probable que las exposiciones hayan cambiado considerablemente en el tiempo (p.e. hábito de fumar y consumo de alcohol), especialmente si la tasa de cambio es diferente en los diferentes grupos. Incluso cuando se dispone de los datos de exposición

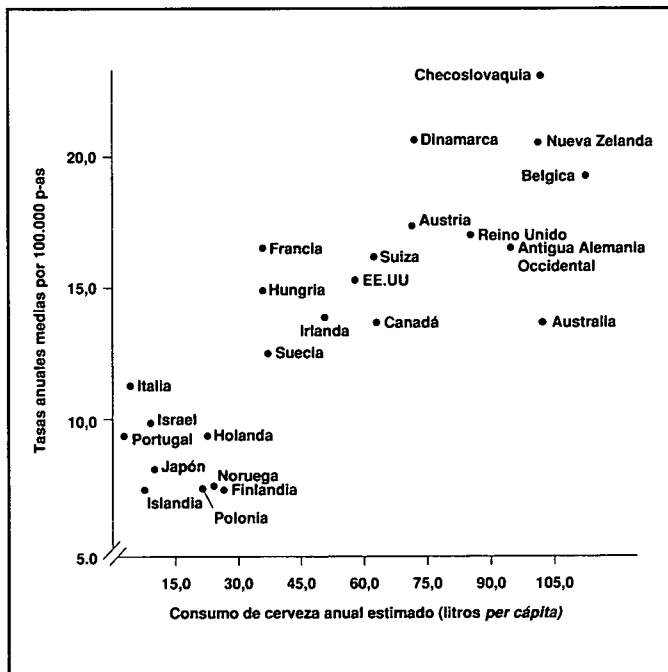


Figura 11.17.

Diagrama de dispersión que muestra la relación entre las tasas estimadas de incidencia medias anuales, ajustadas por la edad para el cáncer rectal entre hombres de 24 países y el consumo de cerveza per cápita, 1960-62 (aprox.) (reproducido con la autorización de Oxford University Press, de Breslow y Enstrom, 1974).

en el pasado, puede que las poblaciones en las que se miden la exposición y la enfermedad no sean las mismas como consecuencia de los cambios dinámicos introducidos por los nacimientos, las muertes y las migraciones.

Una situación en la que no es necesario incorporar un intervalo de tiempo se produce cuando los cambios en la exposición por cohorte de nacimiento están relacionados con cambios en el riesgo de la enfermedad por cohorte de nacimiento o bien cuando los datos de la exposición y la enfermedad se refieren a una cohorte de nacimiento específica. Los Ejemplos 11.16 y 11.17 muestran este punto. El tamaño medio de la familia resumía la experiencia reproductiva de cada generación de mujeres, es decir el número total de hijos al final de su vida reproductiva. Los riesgos de cáncer de ovario de la cohorte utilizada en estos ejemplos también resumen la mortalidad de cada generación de mujeres (para detalles de los cálculos véase Beral *et al.* (1978)). Por tanto, en estas situaciones, no es necesario incorporar un periodo de tiempo de latencia en el análisis.

Conclusiones

A pesar de sus limitaciones, los estudios ecológicos han sido útiles para describir diferencias entre poblaciones. Incluso aunque estén confundidos por factores desconocidos o incontrolables, estas diferencias indican, como mínimo, la presencia de efectos merecedores de más estudios.

Los estudios ecológicos son especialmente adecuados para identificar los factores responsables de las diferencias de riesgo *entre* poblaciones en lugar de las variaciones del riesgo *dentro de* la misma población. Por ejemplo, las comparaciones internacionales han mostrado una fuerte relación entre la ingesta de grasa y el riesgo de cáncer de mama. No obstante, la mayoría de estudios individuales realizados en poblaciones no han podido mostrar esta relación. Se ha propuesto que una posible razón de esta diferencia en los resultados es que la variabilidad entre poblaciones en los niveles de ingesta de grasa es mucho mayor que la variación interindividual dentro de cada población.

Lecturas adicionales

* El libro de Estève *et al.* (1994) proporciona una revisión extensa (aunque elaborada estadísticamente) de los métodos utilizados en el análisis de estudios basados en datos rutinarios.

* Una discusión más completa sobre los estudios ecológicos también puede encontrarse en los artículos de Greenland y Morgenstern (1989), Walter (1991a,b) y Greenland y Robins (1994).

* En Parkin y Khlal (1996) también se incluye una discusión sobre la justificación y metodología de los estudios de emigrantes en epidemiología del cáncer.

Cuadro 11.1. Puntos esenciales

- Los estudios basados en datos rutinarios utilizan sistemas rutinarios de vigilancia para obtener datos sobre la(s) exposición(es) y enfermedad(es) de interés. Este tipo de estudio puede realizarse sin ponerse en contacto directo con ninguno de los sujetos estudiados.
- Los estudios basados en datos rutinarios pueden llevarse a cabo a nivel *individual* (si el individuo es la unidad del estudio) o a un nivel *ecológico* (*agregado*) (si el grupo es la unidad del estudio).
- Las principales ventajas de los estudios basados en datos rutinarios son:

Para estudios a nivel individual y ecológicos

1. Muy económicos y rápidos ya que generalmente utilizan datos existentes sobre la exposición y la enfermedad, sin los costes implicados en la recopilación.
2. Permiten el estudio de gran cantidad de personas y, por tanto, pueden investigarse pequeños aumentos del riesgo.
3. Pueden incluir poblaciones con un rango muy amplio de su nivel de exposición (más de lo que puede encontrarse en una población aislada utilizada para estudios convencionales de cohorte o caso-control).

Para estudios ecológicos

4. Pueden resultar el único método práctico si el nivel de exposición es relativamente homogéneo en una población, pero difiere entre poblaciones (p.e. calidad del agua), o cuando son imposibles las medidas individuales de la exposición (p.e. contaminación del aire).
- Los principales inconvenientes de los estudios basados en datos rutinarios son:

Para estudios a nivel individual y ecológicos

1. El número de variables para las que se dispone de datos es limitado.
2. Es difícil evaluar errores en la medida de las variables de exposición y enfermedad ya que los datos no los recopilan los propios investigadores.
3. Raramente se dispone de los datos sobre variables potenciales de confusión (excepto sexo y edad).

Para estudios ecológicos

4. Susceptibles de la *falacia ecológica*.

Cuadro 11.1. Puntos esenciales (continuación)

5. Es difícil controlar la confusión incluso cuando se dispone de los datos sobre las variables de confusión potenciales.
6. Normalmente es difícil incorporar un periodo de tiempo de latencia adecuado.

Capítulo 12

Introducción al análisis de la supervivencia

12.1 Introducción

En este capítulo examinamos otro método utilizado en el análisis de estudios de intervención, estudios de cohorte y datos recopilados rutinariamente por registros de cáncer hospitalarios y poblacionales. Consideremos el siguiente ejemplo:

Ejemplo 12.1. Una cohorte de 40 mujeres diagnosticadas de cáncer de mama en un determinado hospital durante los años 1989-1993 se siguió desde su diagnóstico hasta finales de 1995 para evaluar su supervivencia. La Tabla 12.1 presenta las fechas de diagnóstico y muerte (o del último contacto) para cada uno de los sujetos del estudio.

En el Ejemplo 12.1, las pacientes se incluyeron y abandonaron el estudio en diferentes momentos en el tiempo (Tabla 12.1). En capítulos anteriores (Capítulos 4, 7-8) se comentó una manera de analizar este tipo de datos que tienen en cuenta las diferentes duraciones individuales del seguimiento. Esta estrategia implicaba el cálculo de tasas basado en unidades de *persona-tiempo a riesgo*. Los cálculos se basaban en la suposición de que la tasa considerada permanezca aproximadamente constante en el tiempo, de modo que 100 personas-años de observación se tratan de manera idéntica, ya incluyan 100 sujetos seguidos durante un año o 50 sujetos seguidos durante dos años.

No obstante, en muchas situaciones, la tasa de aparición del acontecimiento en estudio puede no permanecer constante en el tiempo. Por ejemplo, la probabilidad de morir puede aumentar repentinamente con la aparición de la enfermedad y después disminuir gradualmente al aumentar el intervalo de tiempo desde el diagnóstico. El método más adecuado en estas situaciones es realizar un *análisis de supervivencia*.

12.2. Estimación de la supervivencia

El primer requisito para la estimación de la supervivencia es una *definición clara y concreta del caso*. Para pacientes con cáncer, debe especificarse claramente el tipo de cáncer, la histología, el estadio y el sexo de los pacientes. En el Ejemplo 12.1, se incluyeron en el análisis todos los casos de cáncer de mama confirmados mediante histología.

El segundo requisito es definir de forma clara el momento de inicio del seguimiento. Con frecuencia se utilizan las fechas del primer diagnóstico,

Tabla 12.1.
Seguimiento de 40 mujeres diagnosticadas de cáncer de mama durante los años 1989-1993 en un hospital determinado: datos hipotéticos.

Nº de paciente en estudio	Edad (años)	Estadio ^a	Fecha de diagnóstico	Fecha del último contacto o de muerte	Estado vital en el último contacto ^b	Causa de muerte ^c	Nº de años enteros de observación desde el diagnóstico hasta el último contacto o muerte	Nº de días desde el diagnóstico hasta el último contacto o muerte
1	39	1	01/02/1989	23/10/1992	V	-	3	1.360
2	55	1	22/03/1989	12/02/1995	V	-	5	2.153
3	56	2	16/04/1989	05/09/1989	M	CM	0	142
4	63	1	23/05/1989	20/12/1992	M	CM	3	1.307
5	62	2	12/06/1989	28/12/1995	V	-	6	2.390
6	42	2	05/09/1989	17/12/1990	V	-	1	468
7	45	1	05/10/1989	04/08/1995	V	-	5	2.129
8	38	2	30/11/1989	11/10/1991	M	CM	1	680
9	53	2	07/01/1990	25/10/1990	M	CM	0	291
10	55	1	03/02/1990	31/01/1991	M	CM	0	362
11	49	2	23/03/1990	29/08/1992	V	-	2	890
12	61	1	28/04/1990	13/05/1994	V	-	4	1.476
13	58	1	14/05/1990	01/06/1990	V	-	1	383
14	45	2	15/07/1990	10/09/1993	M	CM	3	1.153
15	60	2	03/08/1990	27/11/1994	V	-	4	1.577
16	69	1	31/08/1990	06/10/1995	M	O	5	1.862
17	58	2	18/09/1990	02/01/1993	M	CM	2	837
18	54	2	09/11/1990	18/06/1995	V	-	4	1.682
19	56	2	28/11/1990	27/06/1995	M	CM	4	1.702
20	52	1	12/12/1990	13/05/1995	M	O	4	1.613
21	67	2	24/01/1991	23/12/1994	M	CM	3	1.429
22	64	2	17/02/1991	06/09/1994	M	O	3	1.297
23	73	1	21/04/1991	24/12/1993	V	-	2	978
24	48	2	09/06/1991	26/06/1994	V	-	3	1.113
25	42	2	20/06/1991	15/03/1992	M	CM	0	269
26	56	2	25/08/1991	19/08/1994	V	-	2	1.090
27	43	1	01/03/1992	06/06/1994	M	CM	2	827
28	64	2	12/04/1992	13/02/1995	M	O	2	1.037
29	35	2	13/04/1992	15/04/1994	M	CM	2	732
30	77	1	05/05/1992	10/05/1995	V	-	3	1.100
31	59	2	10/08/1992	08/11/1992	M	CM	0	90
32	68	1	13/10/1992	21/10/1993	M	CM	1	373
33	70	1	19/11/1992	20/12/1995	V	-	3	1.126
34	58	1	17/01/1993	29/10/1994	V	-	1	650
35	75	2	02/02/1993	10/03/1994	M	CM	1	401
36	55	2	02/05/1993	29/09/1993	M	CM	0	150
37	45	1	11/05/1993	07/02/1994	M	CM	0	272
38	69	1	09/11/1993	26/05/1995	V	-	1	563
39	70	1	07/12/1993	27/05/1995	V	-	1	536
40	27	1	31/12/1993	03/06/1995	V	-	1	519

^a Estadio: 1 = ausencia de ganglios linfáticos regionales afectados y de metástasis
2 = ganglios linfáticos regionales afectados y/o presencia de metástasis regionales o distantes

^b V=viva; M=muerta

^c CM=cáncer de mama; O=otras causas diferentes del cáncer de mama

del inicio del tratamiento o de la admisión al hospital. Aunque la fecha de inicio de la fase clínica de la enfermedad parecería más adecuada, ésta generalmente es difícil de definir. En ensayos clínicos, el punto de inicio apropiado es el tiempo de aleatorización debido a que es éste el momento

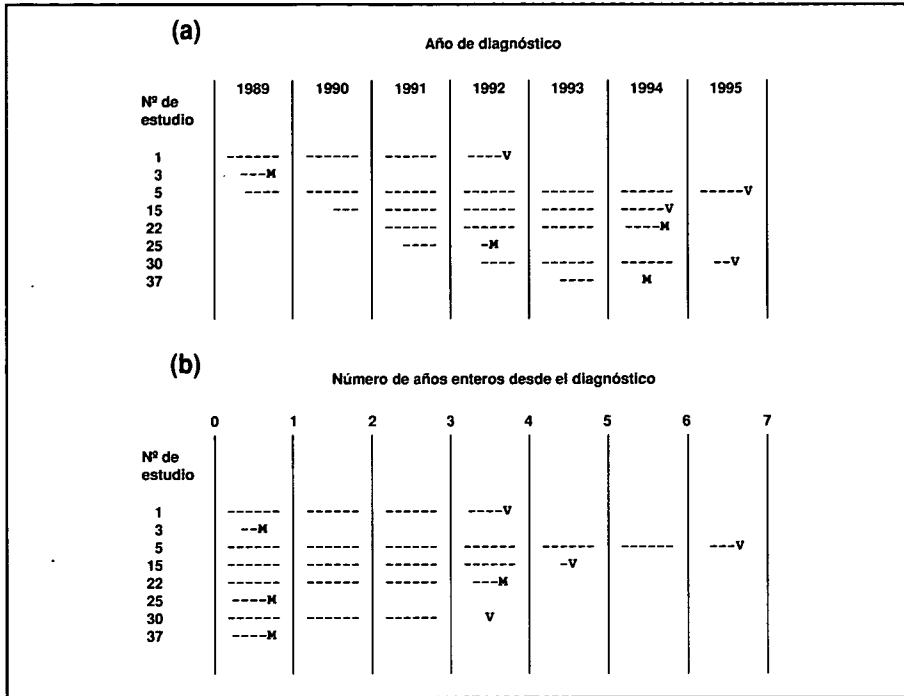


Figura 12.1.

Diagrama que ilustra cómo pueden representarse los datos de seguimiento de 8 de las 40 mujeres con cáncer de mama (véase Tabla 12.1) (a) por año de diagnóstico y (b) por tiempo desde la inclusión en el estudio (V = viva; M = muerta).

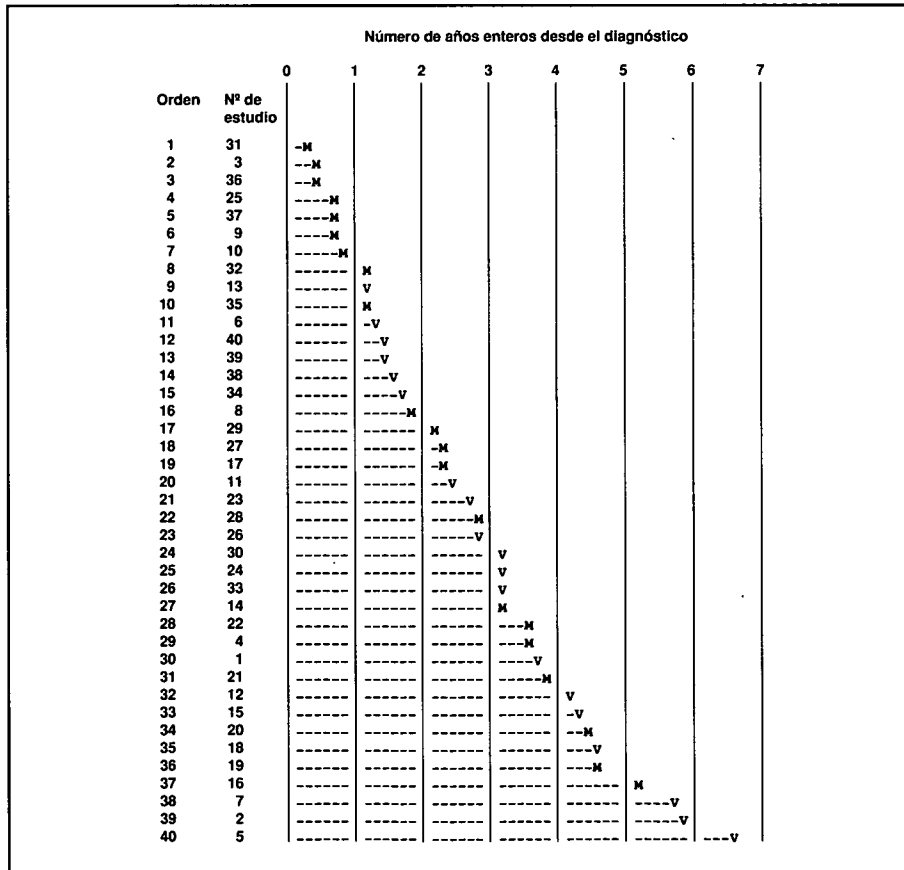


Figura 12.2.

Datos de la Tabla 12.1 ordenados por duración del tiempo de supervivencia observado: (M) representa muerta y (V) viva al final del periodo de seguimiento.

en que los grupos de tratamiento son comparables. En el Ejemplo 12.1, se tomó la fecha de diagnóstico como punto de inicio.

El tercer requisito es definir de forma clara el momento final del seguimiento. La muerte es frecuentemente el punto final de interés, pero no es necesario que sea así. Puede ser también la recidiva de un tumor, la primera aparición de una complicación en particular u otros. El único requisito es que el resultado sea una variable binaria (p.e. estar vivo frente a estar muerto) y que cada sujeto pueda tener solamente un punto final. En nuestro Ejemplo (12.1), la muerte se consideró el punto final de interés.

El tiempo entre el punto de inicio y la aparición del punto final de interés (o la fecha del último contacto) se conoce como *tiempo de supervivencia*. El cálculo del tiempo de supervivencia para algunos de los pacientes en la Tabla 12.1 se presenta en la Figura 12.1(b). Cabe destacar que los sujetos pueden tener fechas de diagnóstico diferentes pero presentar el mismo tiempo de supervivencia. Por ejemplo, las pacientes número 25 y 37 presentaban un tiempo de supervivencia similar a pesar de sus diferentes fechas de inclusión (20/06/1991 y 11/05/1993, respectivamente; Figura 12.1(a); Tabla 12.1). La Figura 12.2 muestra los tiempos de supervivencia individuales para las 40 mujeres con cáncer de mama del Ejemplo 12.1, en orden creciente de duración.

La interpretación de los resultados de un estudio de supervivencia depende en gran parte del tiempo durante el cual se ha seguido cada persona. Un estudio de supervivencia característico implica un periodo de inclusión de pacientes durante el cual se identifican los pacientes y se inicia su seguimiento, un periodo de seguimiento durante el cual los pacientes se siguen pero no se realizan más inclusiones y una fecha de terminación para el análisis. En el Ejemplo 12.1, el periodo de inclusión sólo fue desde principios de 1989 hasta el final de 1993, el periodo de seguimiento continuó desde el principio de 1994 hasta el final de 1995 y la fecha de terminación para este análisis fue el final de 1995.

Una manera de resumir los datos de supervivencia es indicar la proporción de pacientes todavía vivos en un momento determinado en el tiempo. En el Ejemplo 12.1, podemos limitar inicialmente nuestro análisis a pacientes para las cuales disponemos de información completa de los 2 primeros años de seguimiento. La Figura 12.2 muestra que se perdió el seguimiento de 6 mujeres (números de estudio 13, 6, 40, 39, 38 y 34) antes de terminar un periodo de 2 años y por tanto deben excluirse del análisis.

Resumiendo, 34 pacientes terminaron un seguimiento de 2 años, de las cuales 10 murieron y 24 todavía estaban vivas (Figura 12.2). Estos resultados pueden presentarse en un diagrama de ramas, en el cual la rama superior corresponde a las muertes y la rama inferior a las supervivientes (Figura 12.3).

Basándose en estos resultados podemos estimar la *probabilidad (o el riesgo) de morir* en los dos primeros años como $10/34 = 0,29 = 29\%$.

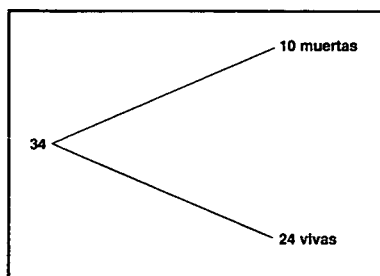


Figura 12.3. Diagrama de ramas que ilustra los dos resultados posibles para las 34 pacientes que terminaron un periodo de seguimiento de 2 años.

12.3 Observaciones censuradas

Un grupo cerrado está formado por un grupo de individuos en el que sólo existen observaciones completas. En la práctica, es raro encontrar un grupo cerrado porque en casi todas las situaciones prácticas habrá algunos sujetos cuyo seguimiento sea incompleto. Esto se producirá bien porque se incorporan a la cohorte demasiado tarde para completar el seguimiento apropiado antes de que termine el estudio o bien porque se pierden durante el seguimiento (p.e. por cambio de dirección, migración). La terminación prematura del seguimiento debida a alguna de estas razones se denomina *censura*.

Nuestro cálculo anterior de la probabilidad de morir en los dos primeros años excluía las observaciones censuradas. No obstante, el hecho de que los sujetos censurados no murieran durante el tiempo que estuvieron en el estudio proporciona alguna información sobre la probabilidad de morir. Supongamos que no sabemos las fechas exactas en las que se produjo censura y todo lo que sabemos es el número de pacientes que no pudieron terminar el periodo de seguimiento definido. Si el intervalo de tiempo es relativamente corto, podemos hacer una estimación simple para suponer que, en promedio, observamos cada paciente censurado durante la mitad del periodo de seguimiento sin observar muertes. Por tanto, para una cohorte de tamaño N con M muertes observadas y P pérdidas debidas a censuras, podemos estimar la probabilidad de morir en el intervalo como

$$M/(N - 0,5 P)$$

Por tanto, la censura reduce el tamaño efectivo de la cohorte a la mitad del tamaño del grupo perdido del seguimiento, es decir, $0,5 P$. Esta forma bastante cruda de considerar la censura funciona correctamente siempre que P sea pequeña en comparación con N .

Ahora podemos volver a calcular la probabilidad de morir en los dos primeros años en el Ejemplo 12.1. De las 40 pacientes con cáncer de mama incluidas en el estudio

- 10 murieron durante el periodo de seguimiento de 2 años ($M = 10$)
- 24 seguían vivas al final del seguimiento ($V = 24$)
- 6 sobrevivieron pero se perdieron durante el seguimiento ($P = 6$)

Estos resultados pueden presentarse en un diagrama de ramas similar al descrito en la Figura 12.3 a excepción de que ahora existe una rama media adicional correspondiente a las observaciones censuradas (Figura 12.4).

Ahora podemos incluir las 40 pacientes en nuestros cálculos. No obstante, el tamaño efectivo de la cohorte no es de 40 sino de 37 debido a las 6 observaciones censuradas ($= 40 - 0,5 \times 6$). La probabilidad de morir se estima como $10 / 37 = 0,27 = 27\%$.

Asimismo, podemos calcular la probabilidad de morir durante los tres primeros años de diagnóstico. Como el último intento de ponerse en con-

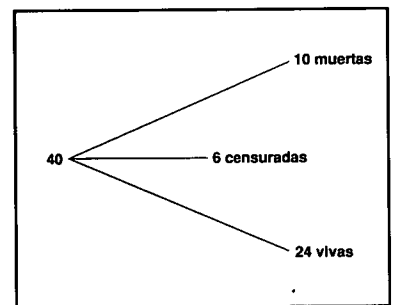


Figura 12.4. Diagrama de ramas que ilustra la evolución de las 40 pacientes de cáncer de mama del Ejemplo 12.1 al final de un periodo de seguimiento de 2 años.

tacto con las pacientes fue en 1995, las pacientes diagnosticadas después del 31 de diciembre de 1992 se incluyeron en el estudio demasiado tarde para poder terminar un seguimiento de 3 años. Por tanto, se censuraron las observaciones de 4 pacientes (números 34, 38-40) (Tabla 12.1). Otras cinco mujeres (números de estudio 13, 6, 11, 23, 26) no terminaron el periodo de observación de 3 años porque se perdieron del seguimiento (p.e. por cambio de dirección) (Figura 12.2). Por tanto, de las 40 pacientes con cáncer de mama incluidas en el estudio

14 murieron durante el periodo de seguimiento de 3 años
 17 todavía estaban vivas al final del seguimiento
 9 se perdieron durante el seguimiento o se incluyeron en la cohorte demasiado tarde para completar los tres años de observación.

A continuación se muestra el diagrama de ramas correspondiente:
 La probabilidad de morir en los tres primeros años después de un diagnóstico de cáncer de mama puede estimarse como $14 / (40 - 0,5 \times 9) = 39\%$.

12.4. Intervalos de seguimiento consecutivos

La utilización de un intervalo de seguimiento único presenta varias limitaciones. En primer lugar, es un modo muy crudo de resumir la supervivencia de una cohorte ya que ignora toda información sobre cuándo se produjeron las muertes y las censuras. Para los cálculos sólo es necesario el número total de muertes y el número total de observaciones censuradas que se produjeron durante el intervalo definido. En segundo lugar, la supervivencia de diferentes cohortes sólo puede compararse si se utilizan los mismos intervalos de seguimiento. Por ejemplo, no puede compararse la supervivencia de dos cohortes de pacientes de cáncer de mama si la supervivencia de una cohorte se resume como la probabilidad de morir en los dos primeros años después del diagnóstico y la supervivencia de la otra, como la probabilidad de morir en los 5 primeros años.

Una manera de evitar estas limitaciones es utilizar unos intervalos de tiempo consecutivos más cortos, en lugar de un único intervalo largo. La supervivencia de la cohorte durante cada uno de estos intervalos puede utilizarse para evaluar la supervivencia durante todo el periodo. En lugar de un único cálculo de la probabilidad de morir habrá una secuencia de cálculos, uno para cada uno de los intervalos.

Volvamos a considerar el seguimiento a 3 años mostrado en la Figura 12.5. Este periodo de 3 años puede dividirse en tres intervalos de un año. Podemos utilizar los datos mostrados en la Figura 12.2 para presentar el número de pacientes que contribuye a cada uno de los tres resultados posibles (es decir, muerte, censura y supervivencia) en cada uno de los tres años consecutivos de seguimiento. El diagrama de ramas resultante se muestra en la Figura 12.6.

En este diagrama de ramas, las supervivientes al final del primer año constituyen la muestra inicial del año siguiente. En el primer año, había 40

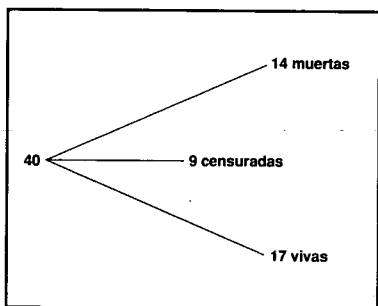


Figura 12.5.
 Diagrama de ramas que ilustra la evolución de las 40 pacientes de cáncer de mama del Ejemplo 12.1 al final de un periodo de seguimiento de 3 años.

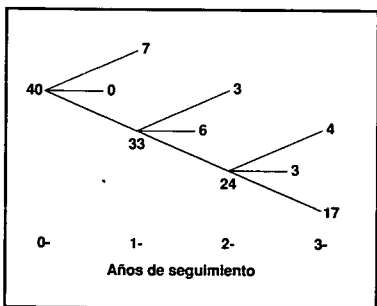


Figura 12.6.
 Diagrama de ramas que muestra el número de pacientes de cáncer de mama del Ejemplo 12.1 que contribuyen a los diferentes resultados en cada uno de los tres primeros años de seguimiento.

pacientes de cáncer de mama siete de las cuales murieron y ninguna se censuró, dejando 33 pacientes al inicio del segundo año. De estas 33 pacientes, tres murieron y seis se censuraron durante el segundo año, dejando 24 al inicio del tercer año. Durante el tercer año, murieron cuatro mujeres, tres se censuraron y 17 estaban vivas.

12.5 Estimación de probabilidades

Ahora podemos sustituir el número de pacientes en nuestros diagramas de ramas por las probabilidades de morir y sobrevivir en cada uno de los intervalos. La probabilidad de morir en cada intervalo puede calcularse como antes, teniendo en cuenta las observaciones censuradas en este intervalo. La *probabilidad de supervivencia* en el intervalo es uno menos la probabilidad de morir en el intervalo.

En el *primer año* no hubo observaciones censuradas, así

- la probabilidad de morir durante el primer año es $7/40 = 0,175$
- la probabilidad de sobrevivir al primer año es $1 - 0,175 = 0,825$.

En el *segundo año* se censuraron seis mujeres. El tamaño efectivo de la cohorte en este año puede estimarse como $33 - (0,5 \times 6) = 30$. Por tanto

- la probabilidad estimada de que un sujeto muera durante el segundo año, considerando que el sujeto estaba vivo al principio del año, es $3/30=0,10$.
- la probabilidad estimada de sobrevivir al segundo año es $1 - 0,10 = 0,90$.

En el *tercer año* se censuraron tres mujeres. El tamaño efectivo de la cohorte es $24 - (0,5 \times 3) = 22,5$. Por tanto

- la probabilidad estimada de que un sujeto muera durante el tercer año, considerando que el sujeto estaba vivo al principio del año, es $4/22,5 = 0,178$.
- la probabilidad estimada de sobrevivir al tercer año es $1 - 0,178 = 0,822$.

En la Figura 12.7 se muestra todo el árbol con las ramas indicando las probabilidades (condicionales) de morir cada año considerando que el sujeto había sobrevivido los años anteriores

Ahora existen cuatro posibles puntos finales de interés, correspondientes a los extremos de las ramas. La probabilidad de cada resultado puede calcularse multiplicando las probabilidades indicadas en cada rama del árbol. Por lo tanto las probabilidades para cada punto final son:

1. Probabilidad de morir durante el primer año = $0,175$
2. Probabilidad de morir durante el segundo año (es decir, probabilidad de sobrevivir el año 1 \times probabilidad de morir el año 2) = $0,825 \times 0,10 = 0,083$

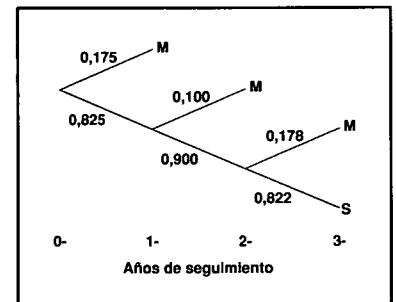


Figura 12.7.

Diagrama de ramas que muestra las probabilidades de cada posible resultado en cada uno de los tres primeros años de seguimiento (M = muerte; S = supervivencia).

3. Probabilidad de morir durante el tercer año = $0,825 \times 0,90 \times 0,178 = 0,132$
4. Probabilidad de estar vivo al final de los tres años = $0,825 \times 0,90 \times 0,822 = 0,610$

Estas probabilidades siempre sumarán 1 ya que no hay otras opciones posibles. La probabilidad de morir en *algún momento* durante el intervalo de 3 años es igual a $0,175 + 0,083 + 0,132 = 0,390$. Esta probabilidad puede calcularse más fácilmente restando de 1 la probabilidad de sobrevivir todo el periodo de 3 años: $1 - 0,610 = 0,390$.

La probabilidad final de sobrevivir (0,610) es un ejemplo de una *probabilidad de supervivencia acumulada* para la cohorte, es decir, la probabilidad de sobrevivir tres años consecutivos.

12.6. Tabla de vida actuarial

Los datos de los cálculos anteriores con frecuencia se presentan en forma de una tabla de vida actuarial, que muestra los números de muertes y censuras que se producen en cada intervalo consecutivo. En la Tabla 12.2 se muestra una tabla de vida para las 40 pacientes de cáncer de mama del Ejemplo 12.1.

Tabla 12.2.

Tabla de vida actuarial para las 40 pacientes de cáncer de mama del Ejemplo 12.1.

Año	Nº al inicio del intervalo (N)	Nº de muertes (M)	Nº de pérdidas (P)	Denominador efectivo (N-0,5 P)	Probabilidad de morir durante el año	Probabilidad de sobrevivir el año	Supervivencia acumulada
0-	40	7	0	40	0,175	0,825	0,825
1-	33	3	6	30,0	0,100	0,900	0,743
2-	24	4	3	22,5	0,178	0,822	0,610
3-	17	4	4	15,0	0,267	0,733	0,447
4-	9	2	3	7,5	0,267	0,733	0,328
5-	4	1	2	3,0	0,333	0,667	0,219
6-	1	0	1	0,5	0,0	1,00	0,219
Total		21	19				

En esta tabla, la probabilidad de morir durante cada año se calcula como $M/(N - 0,5 P)$. Por lo tanto, la probabilidad de sobrevivir el año es igual a $1 - M/(N - 0,5 P)$. La supervivencia acumulada se determina por multiplicación de las probabilidades de supervivencia para cada año consecutivo para obtener las probabilidades acumuladas de sobrevivir 1, 2, ..., 6 años. Por ejemplo, la probabilidad de sobrevivir tres años sin morir es $0,825 \times 0,90 \times 0,822 = 0,610$ (el mismo valor calculado anteriormente). Por lo tanto, la tabla de vida es sólo una manera adecuada de representar estas probabilidades y se obtienen del mismo modo que las del diagrama de ramas. Las tablas de vida son útiles para examinar si la probabilidad de morir cambia con la duración del seguimiento y para presentar de manera concisa las medidas resumen de diferentes intervalos de seguimiento.

Las probabilidades de supervivencia acumulada también pueden repre-

sentarse gráficamente como en la Figura 12.8. Esta representación se denomina *curva de supervivencia*. La curva empieza en 1 (todos los pacientes vivos) y con el tiempo disminuye progresivamente hasta 0 (todos los pacientes han fallecido).

12.7 Método Kaplan-Meier

El método de la tabla de vida actuarial descrito en el apartado anterior (12.6) no requiere información sobre el momento exacto en que se producen las muertes o las censuras. Sólo es necesario conocer el estado vital del sujeto en cada uno de los límites de los intervalos. Si se conocen los *tiempos exactos* de las muertes, las probabilidades de supervivencia pueden estimarse inmediatamente después de cada muerte sin necesidad de agregar los datos en intervalos de un año (o de cualquier otra longitud). Este método de estimación de las probabilidades acumuladas de supervivencia se denomina *método de Kaplan-Meier* y es la estrategia de elección siempre que se disponga de los tiempos en los que se producen cada fallecimiento o cada censura (véase en Estève *et al.* (1994) una descripción completa de los cálculos).

Al igual que la curva de supervivencia de la tabla de vida, las estimaciones de Kaplan-Meier pueden utilizarse para representar las probabilidades de supervivencia acumulada. No obstante, en este caso, la representación tiene forma de *línea escalonada* en lugar de una curva suavizada, ya que la supervivencia acumulada disminuye en el momento exacto en que se produce una muerte y permanece en una meseta entre muertes sucesivas. Por ejemplo, la curva para las 40 pacientes de cáncer de mama del Ejemplo 12.1 mostrada en la Figura 12.9 empieza en 1 y continua horizontalmente hasta la primera muerte (paciente número 31) el día 90; en este momento disminuye en función de la probabilidad de morir estimada. A continuación continua horizontalmente hasta la siguiente muerte (paciente 3) el día 142, y así sucesivamente. La gráfica sólo alcanzará el 0 cuando fallezca la paciente con el tiempo de supervivencia más largo. Si esta paciente todavía está viva al final del periodo de seguimiento, la curva de Kaplan-Meier presenta una meseta que empieza en el tiempo de la última muerte y continua hasta el tiempo de super-

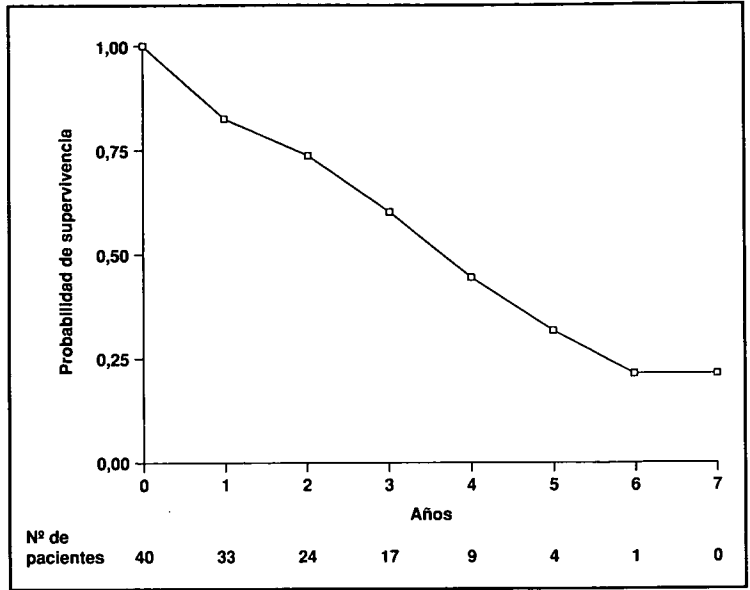


Figura 12.8.

Curva de supervivencia de la tabla de vida (actuarial) para las 40 pacientes de cáncer de mama del Ejemplo 12.1.

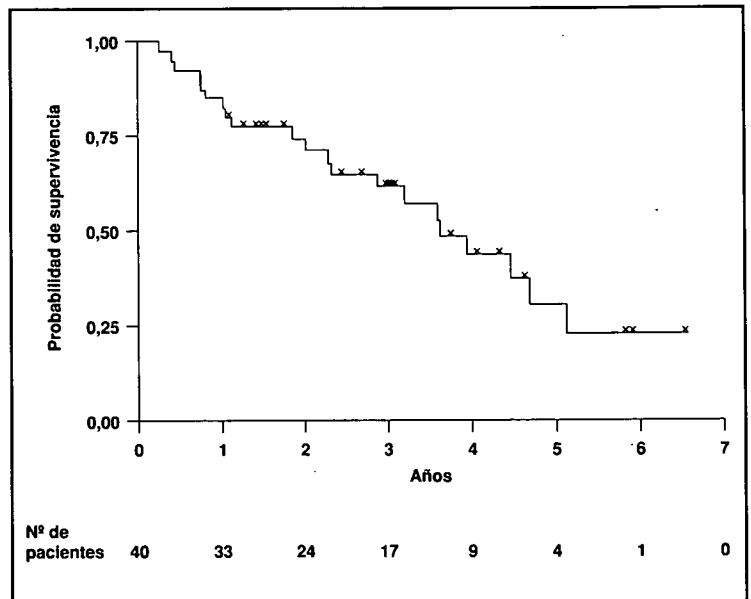


Figura 12.9.

Curva de supervivencia producida por el método Kaplan-Meier para las 40 pacientes de cáncer de mama del Ejemplo 12.1 (x indica los tiempos de censura).

vivencia censurado de la paciente con la mayor supervivencia. En la Figura 12.9, el tiempo de supervivencia de cada observación censurada se indica en la curva con una "x". Tras la última muerte (paciente 16, el día 1.862 (5,1 años)) la curva permanece plana hasta el tiempo de supervivencia censurado más largo (paciente 5, el día 2.390 (6,5 años)).

Es útil indicar el número de pacientes a riesgo en momentos seleccionados en el tiempo (por ejemplo, al inicio de cada año) en la gráfica y/o presentar intervalos de confianza alrededor de las estimaciones de la probabilidad de supervivencia. Esta información es crucial para la interpretación correcta de una curva de supervivencia.

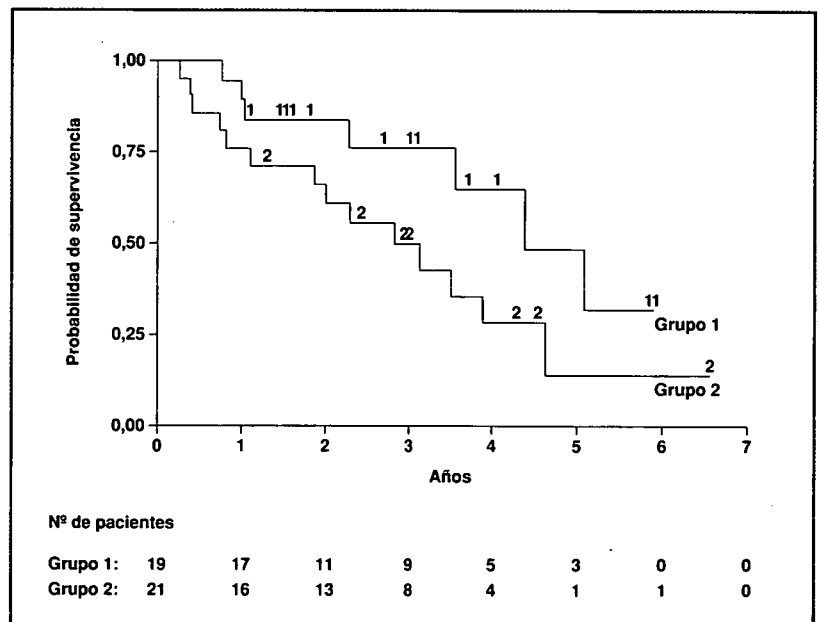
12.8 Comparación de curvas de supervivencia

En muchas situaciones, el objetivo principal del estudio es *comparar* la supervivencia de diferentes grupos de pacientes. Estos grupos pueden definirse según el sexo, estadio del tumor en el momento del diagnóstico (como en el Ejemplo 12.2), tipo histológico, etc. En ensayos clínicos, los grupos se definirán basándose en la asignación del tratamiento. Las probabilidades de supervivencia acumulada pueden calcularse por separado para cada grupo y las dos curvas pueden representarse en el mismo gráfico para su comparación (Figura 12.10).

Ejemplo 12.2. En el Ejemplo 12.1, los investigadores también recopilaron datos sobre el estadio del tumor en el momento del diagnóstico (Tabla 12.1). A continuación se obtuvieron curvas de Kaplan-Meier por separado para cada estadio (Figura 12.10).

Figura 12.10.

Curvas de supervivencia de Kaplan-Meier para pacientes con cáncer de mama según el estadio del tumor en el momento del diagnóstico (grupo 1 = tumor sin afectación de ganglios linfáticos o metástasis; grupo 2 = tumor con afectación de ganglios linfáticos y/o metástasis regionales o distantes). Los números de las curvas de supervivencia representan las observaciones censuradas.



La comparación visual de las curvas de supervivencia es extremadamente útil. Consideremos las gráficas presentadas en la Figura 12.11. En la Gráfica (a), las dos curvas se superponen en los 2 primeros años de seguimiento pero divergen más adelante. En la Gráfica (b), el grupo A presenta inicialmente una supervivencia mejor que el grupo B pero las curvas se cruzan después de 4 años de seguimiento y al final el grupo A tiene una supervivencia peor que el grupo B.

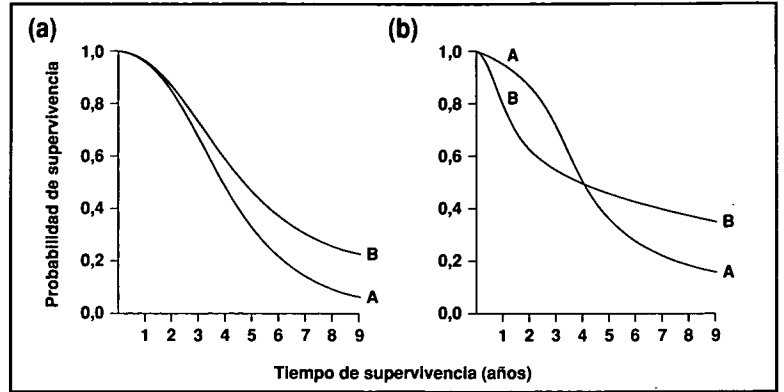


Figura 12.11.

Dos ejemplos posibles de comparación de curvas de supervivencia (reproducidas con la autorización de BMJ Publishing Group, de Gorre, 1981).

Estos patrones no se detectarían si la comparación estuviera limitada a un periodo de seguimiento específico. Por ejemplo, si sólo se calcularan las probabilidades de supervivencia de 2 años, estos indicarían que no había una diferencia clara entre los tratamientos en la Gráfica (a) y que el tratamiento A fue mucho mejor que el tratamiento B en la Gráfica (b). Estos dos ejemplos muestran claramente que la comparación de la supervivencia siempre debe basarse en curvas de supervivencia. Los tests estadísticos para la comparación formal de dos curvas de supervivencia, tales como el *test log-rank*, pueden utilizarse para evaluar la significación estadística de las diferencias observadas (véase Estève *et al.*, 1994).

Al comparar las curvas de supervivencia respecto a un factor pronóstico (o terapéutico) en particular, es importante asegurar que los grupos son similares respecto a otros factores pronósticos. Por ejemplo, en el Ejemplo 12.2, deberían haberse tenido en cuenta otras características tales como edad. Pueden obtenerse curvas de Kaplan-Meier ajustadas por variables de confusión tales como la edad, el sexo, el estadio del tumor, etc. (véase Estève *et al.*, 1994).

12.9 Supervivencia global y supervivencia específica

El primer paso en el análisis de la supervivencia de un grupo de pacientes debe ser examinar su *supervivencia global*. En nuestro ejemplo de cáncer de mama, no se distinguió entre muertes por cáncer de mama y muertes por otras causas. No obstante, un sujeto que fallece en un accidente de tráfico deja de estar a riesgo de morir por cáncer de mama. Una manera de ajustar por estas causas de muerte 'competitivas' es tratar los pacientes que murieron por otras causas como si se hubieran retirado vivos (es decir, censurados en el momento de su fallecimiento) y a continuación efectuar los cálculos de tabla de vida tal y como se ha descrito anteriormente. Las probabilidades de supervivencia obtenidas por este método son *probabilidades de supervivencia específicas*, es decir, tienen en cuenta las muertes debidas a causas diferentes de la enfermedad en estudio.

En el Ejemplo 12.1, cuatro pacientes murieron por causas diferentes del cáncer de mama (véase Tabla 12.1). Para considerar estas muertes como observaciones censuradas puede construirse una nueva tabla de vida actuarial (Tabla 12.3). Como sería de esperar, el número total de muertes disminuyó

Tabla 12.3.

Probabilidad de morir por cáncer de mama calculada mediante el método de la tabla de vida actuarial para las 40 pacientes de cáncer de mama del Ejemplo 12.1. En esta tabla, las muertes por causas diferentes del cáncer de mama se consideraron observaciones censuradas.

Año	N.º al inicio del intervalo (N)	N.º de muertes (M)	N.º de pérdidas (P)	Denominador efectivo (N-0,5 P)	Probabilidad de morir durante el año	Probabilidad de sobrevivir el año	Supervivencia acumulada
0-	40	7	0	40	0,175	0,825	0,825
1-	33	3	6	30,0	0,100	0,900	0,743
2-	24	3	4	22,0	0,136	0,864	0,641
3-	17	3	5	14,5	0,207	0,793	0,509
4-	9	1	4	7,0	0,143	0,857	0,436
5-	4	0	3	2,5	0,0	1,0	0,436
6-	1	0	1	0,5	0,0	1,0	0,436
Total		17	23				

en 4 (17 en lugar de 21) y el número de pérdidas aumentó en 4 (23 en lugar de 19). Asimismo, cuando se conocen las fechas exactas en que se producen las muertes, puede utilizarse el método Kaplan-Meier para estimar estas probabilidades de supervivencia específica (Figura 12.12).

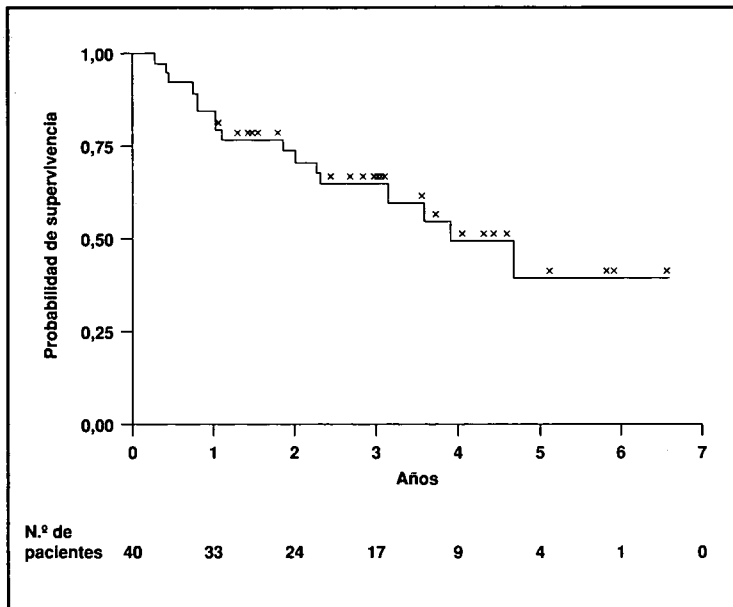
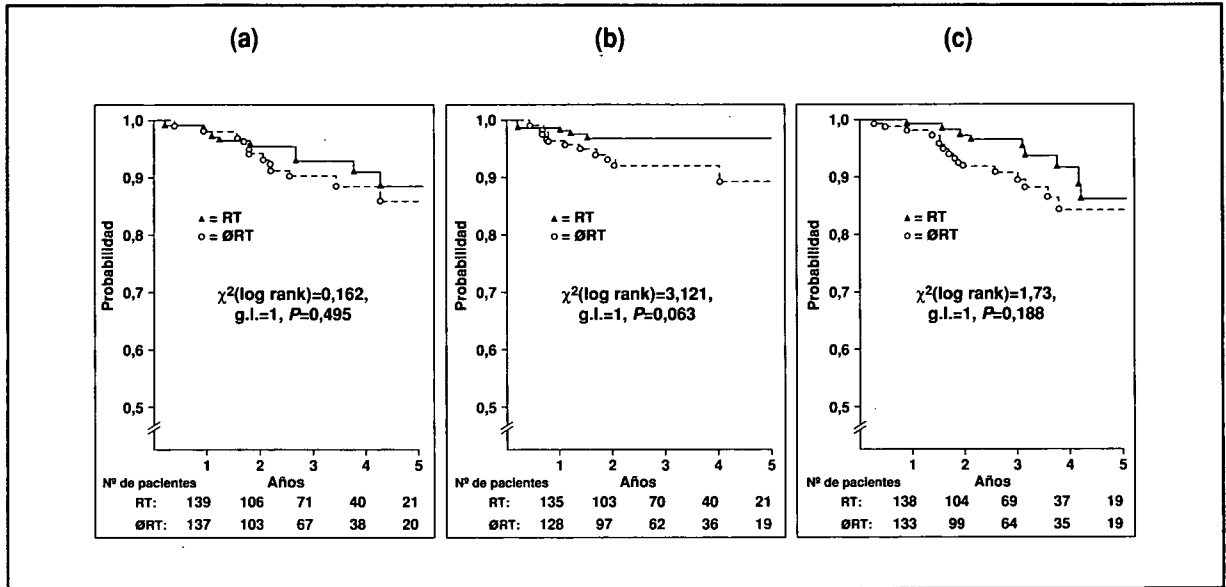


Figura 12.12. Curva de supervivencia específica para las 40 pacientes de cáncer de mama del Ejemplo 12.1, según el método de Kaplan-Meier.

El cálculo de las probabilidades de supervivencia específica requiere información sobre la causa específica de mortalidad. Puede que no sea fácil obtener esta información. Las muertes por otras causas tenderán a estar infranotificadas en pacientes de cáncer debido a que en los certificados de defunción de muchos de ellos se indicará simplemente «muerte por cáncer». Incluso cuando se notifiquen otras causas de muerte diferentes de cáncer, es difícil establecer si la causa de muerte estaba o no relacionada con el cáncer de interés (p.e. cáncer en órganos adyacentes).

Este método no puede utilizarse si no se dispone de datos exactos sobre la causa específica de mortalidad. No obstante, puede compararse la *supervivencia observada* con la *supervivencia esperada* para un grupo de personas de la población general similar al grupo de pacientes, respecto a la raza, sexo, edad y periodo de observación. Esta supervivencia esperada puede obtenerse de las tablas de vida demográficas publicadas rutinariamente (véase más adelante). La comparación proporciona unas *razones de supervivencia relativa* que están ajustadas por la probabilidad de que el paciente muera por una causa diferente de la estudiada (véase en Parkin y Hakulinen (1991) una ilustración de estos cálculos). Por tanto, las razones de supervivencia relativa representan la supervivencia de un grupo de pacientes ajustada a la probabilidad de morir por causas diferentes de la estudiada. En la práctica, las tablas de vida demográficas basadas en la mortalidad por 'todas las causas' y las basadas en 'todas las causas menos cáncer' son muy parecidas y, debido a que las pri-

Ejemplo 12.3. A un total de 381 mujeres con cáncer de mama invasivo en estadio histopatológico I se les sometió a resección local meticulosa del sector mamario afectado además de una disección axilar. Tras esta intervención quirúrgica, 187 se asignaron aleatoriamente para recibir radioterapia postoperatoria en la mama y 194 para no recibir tratamiento adicional. Los eventos finales de interés fueron la supervivencia global y el intervalo de tiempo desde el tratamiento hasta la recidiva local y hasta la aparición de metástasis regionales o distantes. En el análisis se utilizó el método Kaplan-Meier (Uppsala-Örebro Breast Cancer Study Group, 1990). Los principales resultados de este ensayo se muestran en la Figura 12.13.



meras se encuentran más fácilmente, son éstas las que generalmente se utilizan en los cálculos.

12.10 Tablas de vida demográficas

Todos los comentarios anteriores sobre las tablas de vida se refieren a datos obtenidos de cohortes reales, es decir, de grupos de personas que se siguieron realmente en el tiempo.

Las tablas de vida demográficas, calculadas basándose en datos de mortalidad nacionales (o regionales o específicos para un grupo étnico o socio-económico en particular), pueden obtenerse aplicando los riesgos de mortalidad observados actualmente para los diferentes grupos de edad a una cohorte imaginaria. Así, la esperanza de vida al nacer de las mujeres en Inglaterra y Gales, que era de 77 años en 1981 (Bone *et al.*, 1995), depende de la suposición de que las niñas nacidas en 1981 estarán expuestas a los riesgos de morir específicos por edad correspondientes al año 1981 a medida que transcurre su vida (p.e., cuando tengan 30 años en el año 2011, experimentarán los riesgos de mortalidad para los de 30 años observados en 1981). Aunque considerada literalmente esta suposición no es realista, las

Figura 12.13.

Probabilidad de (a) supervivencia global; (b) de permanecer libre de recidiva local; y (c) de permanecer sin metástasis regionales o distantes para 381 pacientes de cáncer de mama según el tipo de tratamiento postoperatorio (RT = radioterapia postoperatoria de mama; Ø RT = ningún tratamiento) (reproducido con la autorización de Oxford University Press, de Uppsala-Örebro Breast Cancer Study Group, 1990).

tablas de vida demográficas son una buena manera de resumir los riesgos actuales de mortalidad. Estas tablas de vida demográficas las preparan y publican normalmente las oficinas de estadísticas gubernamentales.

12.11 Otros eventos finales

Los métodos descritos en este capítulo son parte de un grupo de técnicas estadísticas utilizadas en el 'análisis de supervivencia'. El término 'supervivencia' procede del hecho de que el primer uso de estas técnicas surgió de la industria de seguros, que estaba especialmente interesada en el desarrollo de métodos para estimar primas de seguros. Para hacer esto, necesitaban conocer la esperanza de vida promedio para diferentes tipos de clientes.

No obstante, la utilización de técnicas de análisis de supervivencia no está limitada a estudios en los que la muerte es el evento final de interés. También se ha utilizado ampliamente para estudiar otros resultados tales como fertilidad, cumplimiento con el tratamiento, recidiva de la enfermedad, aparición de complicaciones, etc.

El estudio descrito en el Ejemplo 12.3 tenía más de un desenlace de interés. Los resultados de la Figura 12.13 muestran pocas evidencias de una diferencia en la supervivencia global ($P = 0,5$) o la supervivencia libre de recidiva regional o distante ($P = 0,19$). No obstante, el estudio proporcionó una evidencia moderada a favor de la hipótesis de que las mujeres que recibieron la radiación presentaban un riesgo menor de desarrollar recidivas locales ($P = 0,06$).

12.12 Notas finales

El análisis de la supervivencia puede efectuarse fácilmente con muchos paquetes estadísticos tales como STATA, EGRET, SAS o SPSS. Lamentablemente, no es posible realizar este tipo de análisis en EPI INFO.

La aplicación del análisis de la supervivencia a datos obtenidos por los registros del cáncer se comentará más adelante en el Apartado 17.6.2.

Cuadro 12.1. Puntos esenciales

- El análisis de la supervivencia es otro método utilizado en el análisis de los datos de estudios de intervención, estudios de cohorte y datos recopilados rutinariamente por registros del cáncer. Es especialmente útil cuando la probabilidad de aparición del evento final en estudio cambia con el tiempo desde la inclusión en el estudio.
- La supervivencia de un grupo de personas puede resumirse indicando la proporción de los que todavía están vivos en un punto particular en el tiempo (p.e. al final de un seguimiento de 2 años). No obstante, este método presenta varias limitaciones. En primer lugar, no se tiene en cuenta el tiempo en que se produjeron las muertes y las censuras. En segundo lugar, sólo puede compararse la supervivencia entre grupos de personas si se utiliza el mismo periodo de seguimiento. En tercer lugar, no proporciona evidencia de cambios en la supervivencia con el tiempo de seguimiento.
- Todas las limitaciones anteriores pueden evitarse calculando las *probabilidades de supervivencia acumuladas* para intervalos de seguimiento consecutivos. Estas probabilidades pueden representarse en una gráfica denominada *curva de supervivencia*.
- Las probabilidades de supervivencia acumuladas pueden calcularse utilizando el método de la *tabla de vida actuarial* o el método de Kaplan-Meier. Los dos métodos son esencialmente similares pero la forma de la curva de supervivencia resultante es ligeramente diferente. El método de la tabla de vida actuarial produce una curva suavizada porque las probabilidades de supervivencia acumuladas se calculan sólo al final de cada uno de los intervalos de seguimiento consecutivos mientras que el método de Kaplan-Meier produce una línea escalonada porque estas probabilidades se calculan inmediatamente después de cada muerte.
- El primer paso en el análisis de supervivencia debe ser estimar la *supervivencia global* de toda la cohorte. A veces puede ser útil estimar la *supervivencia específica*. Esto puede realizarse fácilmente si se dispone de información sobre las causas específicas de mortalidad para la cohorte. Si no se dispone de esta información, la supervivencia específica puede todavía estudiarse utilizando la información de tablas de vida demográficas.

Lecturas adicionales

- * El uso de árboles de probabilidad en este capítulo se basó en el método propuesto por Clayton y Hills (1993).
- * En Estève *et al.* (1994) se encuentra una presentación más elaborada de los conceptos estadísticos generales en los que se fundamenta el análisis de supervivencia y su aplicación a los datos recopilados rutinariamente.
- * Peto *et al.* (1976, 1977) presentan una guía para la utilización de las curvas de supervivencia en ensayos clínicos de cáncer.

Capítulo 13

Interpretación de estudios epidemiológicos

En epidemiología, se llevan a cabo estudios para identificar exposiciones que puedan afectar el riesgo de desarrollar una determinada enfermedad u otro acontecimiento relacionado con la salud y para estimar su efecto cuantitativamente. Lamentablemente, los errores son inevitables en casi todos los estudios epidemiológicos, incluso en el ensayo aleatorio mejor realizado. Por tanto, al interpretar los resultados de un estudio epidemiológico, es esencial considerar en qué grado el efecto observado entre una exposición y un resultado pueden estar afectados por errores en el diseño, realización y análisis. Incluso aunque los errores no parezcan una explicación obvia para el efecto observado, antes de concluir que la asociación observada entre exposición y resultado es una relación de causa-efecto real deben plantearse las siguientes preguntas:

- 1) La asociación observada, ¿podría ser debida a errores sistemáticos (*sesgo*) en la forma de selección y seguimiento de los sujetos o en la forma de obtención de la información a partir de ellos?
- 2) ¿Podría ser debida a diferencias entre los grupos en la distribución de otra variable (*de confusión*) que no se midió o consideró en los análisis?
- 3) ¿Podría ser debida al *azar*?
- 4) Por último, ¿es biológicamente plausible que la asociación observada sea *causal*?

La mayoría de estos aspectos se comentó en capítulos anteriores (7-11) en relación a cada diseño de estudio específico. En este capítulo, los consideraremos de manera más estructurada.

13.1 ¿Puede el efecto observado ser debido a un sesgo?

El sesgo tiende a dar lugar a una estimación incorrecta del efecto de una exposición en el desarrollo de una enfermedad u otro resultado de interés. El efecto observado puede estar por encima o por debajo del valor real, dependiendo de la naturaleza del error.

En epidemiología se han identificado muchos tipos de sesgos (Sackett, 1979) pero, para simplificar, pueden agruparse en dos tipos principales: *sesgo de selección* y *sesgo de medida*.

13.1.1 Sesgo de selección

El sesgo de selección se produce cuando existe una diferencia entre las características de las personas seleccionadas para el estudio y las caracterís-

ticas de las personas no seleccionadas. En todos los casos en los que se produce un sesgo de selección, el resultado es una relación entre la exposición y la enfermedad diferente entre los que se incluyeron en el estudio y los que habrían sido elegibles pero no participaron. Por ejemplo, el sesgo de selección se producirá con voluntarios (sesgo de autoselección). Es probable que las personas que se presentan voluntarias a participar en un estudio sean diferentes del resto de la población en una serie de variables demográficas y de estilo de vida (los voluntarios tienden a preocuparse más por la salud, tienden a tener un nivel educativo superior, etc.) algunos de los cuales también pueden ser factores de riesgo para el resultado de interés.

El sesgo de selección puede ser un problema importante en estudios caso-control aunque también puede afectar a estudios transversales y, en menor grado, a estudios de cohorte y ensayos aleatorios.

La selección de una muestra adecuada en un *estudio transversal* no garantiza necesariamente que los participantes sean representativos de la población objeto de estudio ya que puede que algunos de los sujetos seleccionados no participen. Los que no participan pueden introducir un sesgo de selección si difieren de los participantes respecto a los factores en estudio.

En el Ejemplo 13.1, la prevalencia de problemas relacionados con el alcohol aumentó al aumentar la dificultad de reclutar los sujetos, indicando que los que realizaron la entrevista tras muchos intentos de ponerse en contacto con ellos eran diferentes de los que requirieron un menor esfuerzo de reclutamiento, en relación a la variable de interés del estudio. El tiempo y dinero suelen limitar los esfuerzos de reclutamiento a relativamente pocos intentos de contacto. Esto puede sesgar las estimaciones de prevalencia obtenidas del estudio transversal.

En *estudios caso-control*, los controles deben representar a la población de origen de la que se tomaron los casos, es decir, deben proporcionar una estimación de la prevalencia de la exposición en la población general de la que proceden los casos. Esto es relativamente fácil de realizar en un *estudio caso-control anidado* en el que los casos y los controles proceden de una población definida claramente –la cohorte. En un *estudio caso-control poblacional*, también puede definirse una población de origen a partir de la que se obtienen todos los casos (o una muestra aleatoria); los controles se elegirán aleatoriamente a partir de los miembros de la misma población sin la enfermedad.

El método de muestreo utilizado para seleccionar los controles debe asegurar que son una muestra representativa de la población de la que proceden los casos. Si no lo son, se introducirá el sesgo de selección. Por ejemplo, el método utilizado para seleccionar los controles en el Ejemplo 13.2, excluía a mujeres que formaban parte de la población del estudio pero no tenían teléfono. Por tanto, pudo haberse introducido un sesgo de selección de los controles si las mujeres con y sin teléfono difieren respecto a la(s) exposición(es) de interés. Este sesgo puede evitarse excluyendo a los casos que no tenían teléfono, es decir, volviendo a definir la población del estudio como mujeres que viven en hogares con teléfono, de 20-54 años, resi-

Ejemplo 13.1. En St. Louis, Missouri (EE.UU.) se realizó un estudio transversal para evaluar la prevalencia de trastornos psiquiátricos. Este estudio es bastante inusual debido a que se hizo todo lo posible para reclutar el mayor número de sujetos elegibles posible. Los sujetos se eligieron mediante un plan de muestreo de probabilidad de cinco fases que daba a cada familia una probabilidad conocida de ser elegida para el estudio. Una vez seleccionada la familia, se realizó un censo de los residentes por edad y sexo y se eligió aleatoriamente a un residente mayor de 18 años para entrar en el estudio. No se permitió la sustitución. En un 91% de las familias elegibles se realizó un recuento de los residentes. De los 3.778 sujetos seleccionados, 3.004 (80%) fueron entrevistados. La Figura 13.1 muestra que un 32% de los que respondieron se entrevistaron después de dos intentos de contacto y un 66% después de cinco. Hasta el 14º intento no se había terminado un 95% de las entrevistas. El número máximo de intentos que dio lugar a entrevista fue 57. La media para 3.004 que respondieron fue de 5,3; y la mediana 4. Las características demográficas asociadas con una mayor dificultad para ponerse en contacto fueron el ser joven, varón, negro, residente en zona urbana, bien educado y empleado a tiempo completo (Cottler et al., 1987). La Tabla 13.1 muestra las estimaciones de prevalencia de problemas relacionados con el alcohol por número de intentos realizados para obtener una entrevista.

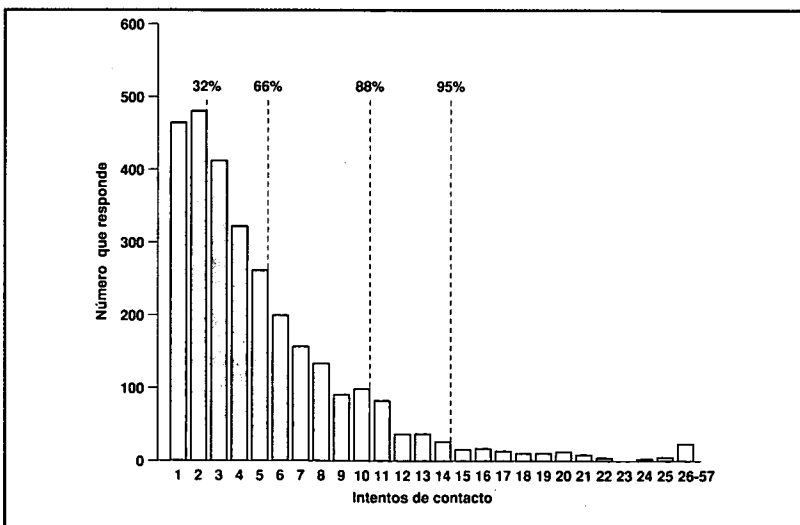
Entrevista terminada en este número de intentos de contacto	Muestra acumulativa (n)	Prevalencia estimada de abuso de alcohol actual y trastornos de dependencia (%)
5	1.943	3,89
7	2.285	3,98
8	2.415	4,22
9	2.511	4,26
57	2.928 ^b	4,61

^a Datos de Cottler et al. (1987).

^b Este número es ligeramente menor que el número de sujetos para los cuales se terminó la entrevista (3.004) debido a que se excluyeron del análisis los cuestionarios en los que faltaban datos.

Tabla 13.1. Estimaciones de la prevalencia de abuso de alcohol actual y trastornos de dependencia según el número de intentos de contacto necesarios para realizar la entrevista, Proyecto del Área Epidemiológica de Captura de St. Louis, 1981-1982^a.

Figura 13.1. Número de intentos de ponerse en contacto necesario para terminar la entrevista. Proyecto del Área Epidemiológica de Captura de St. Louis, 1981-1982 (reproducido con la autorización de Cottler et al., 1987).



Ejemplo 13.2. En el Estudio de Cáncer y Hormonas Esteroides [Cancer and Steroid Hormone Study] se investigó el efecto del uso de anticonceptivos orales sobre el riesgo de los cánceres de mama, endometrio y ovario. La población del estudio consistió en mujeres de 20-54 años residentes en ocho localizaciones geográficas en los Estados Unidos durante el periodo de estudio. Se intentó identificar todos los casos incidentes de cáncer de mama, ovario y endometrio que se produjeron en esta población durante el periodo de estudio, mediante registros del cáncer basados en la población local. Se llamó por teléfono a una muestra aleatoria de hogares, se solicitó información sobre la edad y sexo de todos los miembros de la familia y se seleccionaron los controles entre las mujeres de 20-54 años según unas normas estrictas (Stadel et al., 1985).

dentes en las ocho localizaciones geográficas. Además, el objetivo principal del método de marcado aleatorio de números de teléfono no fue simplemente proporcionar una muestra aleatoria de hogares con teléfono sino una muestra aleatoria de *todas* las mujeres de 20-54 años que vivían en estos hogares durante el periodo de estudio. Esto dependía del grado en que las personas que respondían a los números de teléfono seleccionados estaban dispuestos a proporcionar información exacta sobre la edad y sexo de todos los individuos que vivían en la familia.

Ejemplo 13.3. En Inglaterra, en 1948-52, se realizó un estudio caso-control clásico para examinar la relación entre el hábito de fumar cigarrillos y el cáncer de pulmón. Los casos fueron 1.488 pacientes ingresados por cáncer de pulmón en los hospitales participantes, un 70% de los cuales se encontraban en Londres. Se seleccionó un número similar de controles a partir de los pacientes ingresados en los mismos hospitales por otras enfermedades (a excepción de las enfermedades que en ese momento se suponía estaban relacionadas con el tabaco). Los hábitos de fumar de los controles hospitalarios se compararon con los de una muestra aleatoria de todos los residentes en Londres (Tabla 13.2). Estas comparaciones mostraron que, entre hombres de edades similares, había más fumadores en los controles hospitalarios que en la muestra de la población (Doll y Hill, 1952).

Tabla 13.2.

Distribución, ajustada por edad, del hábito de fumar en los hombres seleccionados como controles hospitalarios sin cáncer de pulmón y en una muestra aleatoria de hombres de la población general de la que proceden los casos^a.

Sujetos	Porcentaje de sujetos					Número de entrevistados (%)
	No fumadores	Cantidad más reciente de cigarrillos fumados al día				
		1-4	5-14	15-24	25+	
Controles hospitalarios	7,0	4,2	43,3	32,1	13,4	1.390 (100)
Muestra de la población general	12,1	7,0	44,2	28,1	8,5	199 (100)

^a Datos de Doll & Hill (1952).

A veces, no puede definirse con precisión la población de la que proceden los casos. En estas circunstancias, pueden utilizarse *controles hospitalarios* por-

que la población de origen puede volver a definirse como 'usuarios de hospitales'. Los controles hospitalarios también pueden ser preferibles por razones logísticas (más fáciles y baratos de identificar y reclutar) y porque presentan un menor potencial de sesgo de memoria (véase más adelante). Pero el sesgo de selección puede introducirse si la admisión al hospital por otros trastornos está relacionada con la exposición de interés de nuestro estudio.

La Tabla 13.2 muestra que entre hombres de edades similares, el porcentaje de no fumadores fue del 7,0% en los controles hospitalarios y del 12,1% en la muestra poblacional. El porcentaje de fumar como mínimo 25 cigarrillos al día fue del 13,4% en los controles hospitalarios pero sólo del 8,5% en la muestra poblacional. Los investigadores indican que esta diferencia en los hábitos de fumar entre los controles hospitalarios y la muestra aleatoria de la población podría explicarse por asociaciones no conocidas anteriormente entre el fumar y algunas enfermedades. Así, la fuerza de la asociación entre el cáncer de pulmón y el fumar estaba *subestimada* en el estudio caso-control debido a que la prevalencia de fumar en los controles de hospitales era mayor que en la población general de la que se tomaron los casos de cáncer de pulmón.

Un grupo de controles hospitalarios puede no reflejar a la población a riesgo debido a que incluye personas ingresadas en el hospital por enfermedades causadas (o evitadas) por las exposiciones de interés. Con objeto de eliminar este tipo de sesgo de selección, deben excluirse los individuos hospitalizados por enfermedades relacionadas con la exposición en estudio. No obstante, esta exclusión no debe ampliarse a los pacientes de hospitales con historia de enfermedades relacionadas con la exposición ya que no se impone una limitación así en los casos. Por tanto, los pacientes ingresados en el hospital debido a un trastorno relacionado con el fumar (p.e. reactivación de una bronquitis crónica) deben excluirse del grupo control en un estudio caso-control sobre la relación entre fumar y cáncer de pulmón mientras que los ingresados por otras condiciones (p.e. accidentes) pero con una historia de bronquitis crónica deben incluirse.

El sesgo de selección es un problema menor en *estudios de cohorte* porque la inclusión de los individuos expuestos y no expuestos se efectúa antes del desarrollo de cualquier resultado de interés. Esto también procede en *estudios de cohorte históricos* ya que la determinación del estado de la exposición se realizó en el pasado, antes de conocer el resultado. No obstante, todavía puede introducirse un sesgo en la selección del grupo de 'no expuestos'. Por ejemplo, en estudios de cohorte ocupacionales en los que la población general se utiliza como grupo de comparación, se suele observar que la morbilidad global de los trabajadores es menor que la de la población general. Esto es debido a que sólo pueden seguir trabajando las personas relativamente sanas, mientras que la población general está formada por una mayor cantidad de personas incluyendo aquellas que están demasiado enfermas como para trabajar. Este tipo de sesgo de selección se denomina *efecto del trabajador sano*. Puede reducirse al mínimo limitando el análisis a personas de la misma fábrica sometidas al mismo proceso de selección pero con un trabajo diferente (véase Apartado 8.2.2.).

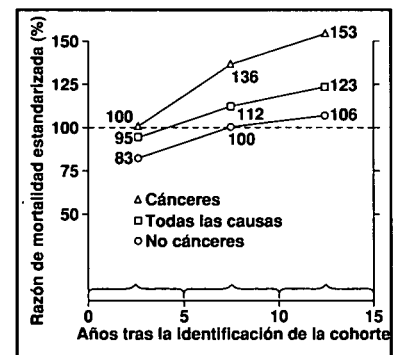


Figura 13.2. Disminución del efecto del trabajador sano con el transcurso del tiempo tras la identificación inicial de una cohorte de trabajadores activos. Gráfico basado en los datos de mortalidad de una cohorte de trabajadores de productos de amianto (Enterline, 1965) (reproducido con la autorización de McMichael, 1976. © American College of Occupational and Environmental Medicine, 1976).

Ejemplo 13.4. *Supongamos que una cohorte en particular estuviera constituida por trabajadores en activo en un momento particular en el tiempo y que después de un periodo de inclusión de sujetos fijo no se permitiese incluir más miembros en la cohorte. Si se siguen todos los miembros de la cohorte incluyendo aquellos que abandonan o se retiran por razones de salud, el efecto del trabajador sano disminuirá con el tiempo porque el proceso inicial de selección de sanos en el momento de inclusión en el grupo de trabajadores se debilitará con el paso del tiempo. Cabe observar que las personas que trabajan no sólo suelen ser más sanas, sino que también, si no lo son, es menos probable que padezcan enfermedades a largo plazo y fáciles de detectar (p.e. enfermedades cardiovasculares) que enfermedades que permanecen asintomáticas durante largos periodos de tiempo (p.e. muchos cánceres) (McMichael, 1976).*

El Ejemplo 13.4 muestra que el sesgo de selección del trabajador sano varía con el tipo de enfermedad, siendo menos marcada para el cáncer que para enfermedades distintas al cáncer (principalmente alteraciones cardiovasculares) y tiende a disminuir con el tiempo desde la inclusión en el grupo de trabajadores.

El seguimiento incompleto debido a la falta de respuesta y negativa a participar también puede ser una fuente importante de sesgo de selección en estudios de cohorte en los que las personas se tengan que seguir durante periodos prolongados de tiempo. No obstante, esto sólo introducirá sesgo si el grado de pérdida es diferente para categorías de exposición diferentes. Por ejemplo, los sujetos pueden estar más dispuestos a volver a la visita de seguimiento si han desarrollado síntomas de la enfermedad. Esta tendencia puede ser diferente en los expuestos y no expuestos, ocasionando una estimación superior (o subestimación) del efecto. Las definiciones de los periodos de seguimiento individuales a veces pueden ocultar esta fuente de sesgo. Por ejemplo, si los sujetos con una ocupación determinada tienden a dejar su trabajo cuando desarrollan síntomas de la enfermedad, puede que los casos expuestos no se identifiquen si el seguimiento termina en el momento en que los sujetos cambian a otro trabajo. En estudios de inmigrantes puede producirse un sesgo de selección similar si las personas que enferman vuelven a sus países de origen antes de que se diagnostique correctamente su enfermedad en el país huésped.

Es menos probable que los *ensayos de intervención aleatorizados* estén afectados por un sesgo de selección, ya que los sujetos se distribuyen aleatoriamente a alguno de los grupos de intervención a comparar. No obstante, la negación a participar tras la distribución aleatoria y las bajas del estudio pueden afectar los resultados si están relacionadas con el estado de exposición. Para reducir al mínimo el sesgo de selección, la asignación a los diferentes grupos de estudio se debe realizar sólo después de haber evaluado que los sujetos sean elegibles y estén dispuestos a participar (véase Apartado 7.9). Los datos también se deben analizar según «intención de tratar» independientemente de que los sujetos cumplan o no con la intervención asignada (véase Apartado 7.12).

13.1.2 Sesgo de medida

El *sesgo de medida* (o *información*) se produce cuando las medidas o clasificaciones de la enfermedad o la exposición no son válidas (es decir, cuando no miden correctamente lo que se supone tienen que medir). Los errores en la medida pueden introducirse por el observador (*sesgo del observador*), por el individuo del estudio (*sesgo del respondedor*), o por los instrumentos (p.e. cuestionario o esfigomanómetro o pruebas de laboratorio) utilizados para realizar las determinaciones (véase Capítulo 2).

Clasificación errónea de un atributo único (exposición o resultado) y prevalencia observada

Supongamos que se realiza un estudio transversal para evaluar la prevalencia de un atributo determinado en una población en estudio determinada. Se administró un cuestionario a todos los participantes elegibles; no hubo rechazos. Denominaremos p^* a la proporción observada en la población que presenta el atributo y p a la prevalencia real. p^* tiene dos componentes. Un componente procede de los individuos con el atributo clasificados por el cuestionario correctamente con el atributo (verdaderos positivos). El otro componente procede de los individuos que en realidad no presentan el atributo pero se han clasificado erróneamente con el atributo (falsos positivos). La proporción de individuos clasificados por el cuestionario con el atributo (p^*) es:

$$p^* = p \times \text{sensibilidad} + (1 - p) \times (1 - \text{especificidad})$$

y por tanto depende de la prevalencia real del atributo (p) y de la sensibilidad y especificidad del método de determinación. Por ejemplo, si $p=0,1\%$, sensibilidad=90% y especificidad=90%, entonces $p^*=10,1\%$. Esto significa que si la prevalencia se estimara por la proporción que es clasificada por el cuestionario como que presentan el atributo, la estimación sería del 10,1% mientras que el valor real sería sólo del 0,1%. Esta clasificación errónea corresponde a una estimación 100 veces superior a la real.

Puede obtenerse una estimación corregida de la prevalencia real en aquellas situaciones en las que se conocen la sensibilidad y la especificidad del método de determinación. La prevalencia también puede estimarse reagrupando la fórmula anterior como sigue:

$$p = \frac{p^* + \text{especificidad} - 1}{\text{sensibilidad} + \text{especificidad} - 1}$$

La Tabla 13.3 muestra el efecto de diferentes niveles de sensibilidad y especificidad de un método de determinación sobre la prevalencia observada de un atributo particular. En la mayoría de los casos, la prevalencia observada es una sobreestimación de la prevalencia real. Por ejemplo, la prevalencia observada al 90% de sensibilidad y 90% de especificidad suponiendo una prevalencia real del 0,1%, es sólo aproximadamente la mitad de la que se obtendría si la prevalencia real fuera del 12,5% (e.d., 10% frente al 20%).

El sesgo en la sobreestimación está muy influenciado por pérdidas de la especificidad (especialmente cuando la prevalencia real es inferior al 50%). Por el contrario, las pérdidas de la sensibilidad muestran, como mucho, unos efectos moderados en la prevalencias observada.

Tabla 13.3. Prevalencia observada (p^*) [%] de un atributo en una población para diferentes niveles de la prevalencia real (p) y diferentes valores de sensibilidad y especificidad del método de determinación.

Prevalencia real (%)	Especificidad (%)	Sensibilidad (%)		
		99	90	80
0,1	99	1,1	1,1	1,1
	90	10,1	10,1	10,1
	80	20,1	20,1	20,1
0,5	99	1,5	1,4	1,4
	90	10,4	10,4	10,4
	80	20,4	20,4	20,3
2,5	99	3,5	3,2	3,0
	90	12,2	12,0	11,8
	80	22,0	21,8	21,5
12,5	99	13,3	12,1	10,9
	90	21,1	20,0	18,8
	80	29,9	28,8	27,5
50,0	99	50,0	45,5	40,5
	90	54,5	50,0	45,0
	80	59,5	55,0	50,0
62,5	99	62,3	56,6	50,4
	90	65,6	60,0	53,8
	80	69,4	63,8	57,5

Clasificación errónea y relaciones exposición-enfermedad

Dos tipos principales de clasificación errónea pueden afectar la interpretación de la relación exposición-enfermedad: no diferencial y diferencial. La *clasificación errónea no diferencial* se produce cuando una exposición o enfermedad no se clasifica correctamente para proporciones iguales de sujetos en los grupos comparados. En otras palabras, la clasificación errónea no diferencial se refiere a errores en la categorización de la enfermedad que no están relacionados con el estado de la exposición. En estas circunstancias, todos los individuos, independientemente de su estado exposición/enfermedad, presentan la misma probabilidad de ser clasificados erróneamente. Por el contrario, la *clasificación errónea diferencial* se produce cuando los errores en la clasificación de la enfermedad dependen del estado de la exposición o vice-versa, cuando los errores en la clasificación de la exposición dependen de la enfermedad.

Estos dos tipos de clasificación errónea afectan a los resultados del estudio epidemiológico de diferente forma. Sus implicaciones también dependen de si la clasificación errónea está relacionada con la exposición o con la enfermedad.

Clasificación errónea no diferencial de la exposición

La clasificación errónea no diferencial de la exposición se produce cuando todos los individuos (independientemente de su estado de salud / enfermedad, actual o futuro) presentan la misma probabilidad de ser clasificados erróneamente respecto a su exposición. Normalmente, este tipo de clasificación errónea ocasiona una subestimación de la fuerza de la asociación entre exposición y enfermedad, es decir, 'diluye' el efecto.

Ejemplo 13.5. Se realizó un estudio caso-control para evaluar si el consumo de café aumentaba el riesgo de cáncer de páncreas. Los sujetos se clasificaron como bebedores «alguna vez» o «nunca». Los resultados reales de este estudio en ausencia de clasificación errónea de la exposición (e.d., método de determinación de la exposición con sensibilidad = 100%; especificidad = 100%) se muestran en la Tabla 13.4.

	Consumo de café		Total
	Alguna vez	Nunca	
Casos de cáncer de páncreas	200	100	300
Controles	150	150	300
Total	350	250	600

Odds ratio real = $(200/100) / (150/150) = 2,00$.

Supongamos que la información sobre el consumo de café se obtuvo mediante un cuestionario autoadministrado y que sólo un 80% de los sujetos del estudio que bebían café normalmente (independientemente de si tenían cáncer de páncreas o no) lo indicaron en el cuestionario (sensibilidad=80%). Asimismo, sólo un 90% de los que nunca bebían café lo mencionaron correctamente en el cuestionario (especificidad=90%).

	Consumo de café		Total
	Alguna vez	Nunca	
Casos de cáncer de páncreas	$(200 \times 0,8) + (100 \times 0,1) = 170$	$(100 \times 0,9) + (200 \times 0,2) = 130$	300
Controles	$(150 \times 0,8) + (150 \times 0,1) = 135$	$(150 \times 0,9) + (150 \times 0,2) = 165$	300
Total	305	295	600

Odds ratio observada = $(170/130) / (135/165) = 1,60$

Tabla 13.4.

Resultados de un estudio caso-control hipotético: estado real de exposición (sensibilidad = 100% y especificidad = 100% para casos y controles).

Tabla 13.5.

Sesgo debido a la clasificación errónea no diferencial de la exposición: estado de exposición observado (sensibilidad = 80% y especificidad = 90% para casos y controles).

En el Ejemplo 13.5, la clasificación errónea no diferencial de la exposición introdujo un sesgo hacia una subestimación del efecto real de la exposición.

La clasificación errónea no diferencial de la exposición también afecta a los resultados de otros diseños de estudio. Por ejemplo, los estudios de cohorte histórica de ocupación se basan en información obtenida retrospectivamente a partir de archivos existentes sobre exposición. No obstante,

en muchos casos no existen medidas específicas históricas de la exposición medioambiental que permitan la clasificación exacta de los individuos. Esto significa que un individuo se clasifica como 'expuesto' o 'no expuesto' por su trabajo o por la pertenencia a una asociación profesional. Estas variables aproximadas son sólo marcadores crudos de los niveles reales de la exposición y su validez es limitada. No obstante, es poco probable que la validez de estos archivos sea diferente para los que desarrollaron años más tarde la enfermedad de interés y para los que no la desarrollaron. Mientras la clasificación errónea sea no diferencial, normalmente diluirá la asociación real entre la exposición y la enfermedad.

Las implicaciones de la clasificación errónea no diferencial de la exposición dependen en gran parte de si el estudio se percibe como 'positivo' o 'negativo'. Este sesgo es un aspecto mayor a considerar en la interpretación de estudios que parecen indicar la ausencia de un efecto. En estas circunstancias, es decisivo para los investigadores considerar el problema de la clasificación errónea no diferencial de la exposición para determinar en qué grado puede haberse dejado de observar un efecto real. Por otro lado, no es correcto negar resultados que indican la existencia de una asociación simplemente porque exista una clasificación errónea no diferencial considerable. En términos generales, la estimación del efecto sin errores de clasificación es incluso mayor que la observada.

Cabe destacar que la clasificación errónea no diferencial de la exposición da lugar a una subestimación del riesgo relativo (sesgo hacia el efecto relativo de 1) sólo si la exposición se clasifica como una variable binaria (p.e. 'expuestos' frente a 'no expuestos').

Cuando existen tres o más categorías de exposición, la clasificación errónea de la exposición no diferencial puede dar lugar a una sobreestimación o a una subestimación del efecto (Flegal *et al.*, 1986). Por ejemplo, al estudiar el efecto de diferentes niveles de consumo de café (clasificados como 'nunca', 'bajo' y 'alto') sobre el riesgo de cáncer de páncreas, el riesgo relativo asociado con niveles bajos de consumo de café estará sobreestimado si existe una tendencia general de los sujetos con un consumo alto real a indicar un menor consumo (pero no tan extremo como para considerarse no bebedores). Por otro lado, si algunas personas con un consumo alto real se clasifican como nunca bebedores, el riesgo relativo asociado con niveles bajos de consumo de café estaría subestimado o incluso podría invertirse (como en el Ejemplo 13.6).

En esta discusión, hemos supuesto que la clasificación errónea se produce entre 'exposición' y 'no exposición' o entre diferentes niveles de exposición. Obviamente, una exposición puede clasificarse erróneamente como otra. Por ejemplo, al estudiar el efecto de anticonceptivos orales sobre el riesgo de cáncer de mama, las mujeres entrevistadas pueden confundir los anticonceptivos con otros fármacos que pueden haber tomado en el pasado.

Ejemplo 13.6. La Tabla 13.6 muestra los resultados de un estudio caso-control hipotético con una tendencia exposición-respuesta positiva. Supongamos que, tanto casos como controles, se clasifican correctamente respecto a la exposición en estudio a excepción de que un 60% de los sujetos que realmente estaban en el grupo de 'ninguna' exposición se clasificaron erróneamente en el grupo de exposición 'alta', y un 60% de los que realmente estaban en el grupo de 'alta', se clasificaron erróneamente en el grupo de 'ninguna' exposición. Mientras las odds ratios en los datos originales fueron 2,00 y 8,00 para las categorías de exposición baja y alta, respectivamente, los valores equivalentes fueron de 0,90 y 0,73 en los datos clasificados erróneamente. La clasificación errónea dio lugar a la creación de una tendencia exposición-respuesta inversa.

Estado de enfermedad	Estado de exposición verdadero		
	Ninguna ^b	Baja	Alta
Casos	53	40	60
Controles	424	160	60
Odds ratio real	1,00	2,00	8,00
Estado de exposición erróneo (60% clasificación errónea)			
	Ninguna ^b	Baja	Alta
Casos	$53 - (53 \times 0,6) + (60 \times 0,6) = 57$	40	$60 - (60 \times 0,6) + (53 \times 0,6) = 56$
Controles	$424 - (424 \times 0,6) + (60 \times 0,6) = 206$	160	$60 - (60 \times 0,6) + (424 \times 0,6) = 278$
Odds ratio observada	1,00	0,90	0,73

^a Datos de Dosemeci *et al.* (1990).
^b Considerada como la categoría de referencia.

Tabla 13.6.

Clasificación errónea no diferencial de una exposición con más de dos categorías^a.

Clasificación errónea diferencial de la exposición

Este tipo de clasificación errónea puede *sesgar las estimaciones de la asociación en las dos direcciones* y, por tanto, puede ser responsable de observar asociaciones falsas.

En el Ejemplo 13.7, era más probable que las pacientes con cáncer de mama indicaran incorrectamente que habían usado alguna vez anticonceptivos orales que los controles sanos, ocasionando una asociación falsa entre anticonceptivos orales y cáncer de mama. Éste es un tipo particular de clasificación errónea diferencial denominado *sesgo de memoria*. Una forma de reducir el sesgo de memoria es utilizando controles hospitalarios porque los controles generalmente tendrán el mismo incentivo que los casos para recordar los acontecimientos en el pasado. Los sujetos también deben mantenerse ignorantes respecto a la hipótesis específica del estudio.

Debido a que no existen evidencias biológicas que indiquen que el efecto de la vasectomía puede ser diferente en hombres católicos y protestantes, la explicación más probable de las observaciones del Ejemplo 13.8 es que los controles católicos habían infranotificado la vasectomía, mientras que los católicos que tenían cáncer de testículo presentaban una menor

tendencia a infranotificar la vasectomía. Esta clasificación errónea de la exposición diferencial sesgó el riesgo relativo hacia valores elevados en hombres católicos.

Ejemplo 13.7. Consideremos el ejemplo de un estudio caso-control sobre el uso de anticonceptivos orales y cáncer de mama. Los resultados reales de este estudio en ausencia de clasificación errónea se muestran en la Tabla 13.7.

Tabla 13.7.

Datos de un estudio caso-control hipotético: estado de exposición real (sensibilidad = 100% y especificidad = 100% para casos y controles).

	<u>Uso de anticonceptivos orales</u>		Total
	Alguna vez	Nunca	
Casos de cáncer de mama	150	200	350
Controles	150	200	350
Total	300	400	700

Odds ratio real = (150/200) / (150/200) = 1,00

Supongamos que un 20% de los casos que nunca habían utilizado anticonceptivos orales indican incorrectamente que los han usado, mientras que todas las usuarias indican correctamente sus costumbres (sensibilidad=100%; especificidad=80%). Los controles indican correctamente su uso (sensibilidad=100%; especificidad=100%).

Tabla 13.8.

Sesgo debido a la clasificación errónea diferencial de la exposición: estado de exposición observado (sensibilidad = 100% y especificidad = 80% en casos; sensibilidad = 100% y especificidad = 100% en controles).

	<u>Uso de anticonceptivos orales</u>		Total
	Alguna vez	Nunca	
Casos de cáncer de mama	$150 + (200 \times 0,20) = 190$	$200 - (200 \times 0,20) = 160$	350
Controles	150	200	350
Total	340	360	700

Odds ratio observada = (190/160) / (150/200) = 1,58

A veces, puede obtenerse una evidencia directa de la presencia y magnitud de la clasificación errónea diferencial como en el Ejemplo 13.9.

En el Ejemplo 13.9, los casos, pero no los controles, comunicaron más frecuentemente su incapacidad de broncearse tras haber sido diagnosticados de melanoma cutáneo –la proporción de casos que indicó incapacidad a broncearse fue del 26% (=9/34) en 1982, antes del diagnóstico de cáncer de piel y del 44% (=15/34) después del diagnóstico.

Asimismo, los observadores que conocen el estado de la exposición de un individuo pueden estar consciente o inconscientemente predispuestos a evaluar las variables del resultado según la hipótesis en estudio. Este tipo de sesgo se conoce como *sesgo del observador*. Por ejemplo, en un ensayo para evaluar si un fármaco disminuye la presión arterial, los observadores pueden subestimar la presión arterial en el grupo de tratamiento o sobreestimar la presión arterial en el grupo control. Una manera de reducir este tipo de sesgo es mantener a los observadores “ciegos” en relación al estado de

Ejemplo 13.8. En el oeste del estado de Washington se realizó un estudio caso-control para evaluar el efecto de la vasectomía (esterilización quirúrgica masculina) sobre el riesgo de cáncer de testículo. La información sobre la exposición (vasectomía) se obtuvo por entrevistas telefónicas. Los resultados mostraron un 50 por ciento de riesgo en exceso asociado a la vasectomía: odds ratio (OR) = 1,5 (IC 95%: 1,0-2,2). No obstante, más análisis mostraron que el efecto de la vasectomía era considerablemente diferente para hombres católicos y protestantes: OR = 8,7 para católicos y OR = 1,0 para protestantes. Mientras los casos católicos y no católicos indicaron una historia de vasectomía con aproximadamente la misma frecuencia, sólo un 6,2% de los controles católicos comunicó una historia de vasectomía frente a un 19,7% de otros controles (Strader et al., 1988).

Religión	Casos		Controles		Odds ratio ajustada por la edad (IC 95%) ^c
	N. ^o	% con vasectomía ^b	N. ^o	% con vasectomía	
Protestantes	129	18,0	295	19,0	1,0 (0,6-1,7)
Católicos	42	24,4	96	6,2	8,7 (2,8-27,1)
Otros	57	32,3	122	21,0	1,3 (0,6-3,0)

^a Datos de Strader et al. (1988).
^b Porcentaje ajustado a la distribución de edad de todos los controles.
^c IC 95% = intervalo de confianza al 95%.

Tabla 13.9.

Historia de vasectomía en casos de cáncer de testículo y controles según religión^a.

Ejemplo 13.9. El Estudio de salud de enfermeras (The Nurser's Health Study) es un estudio de una cohorte de 121.700 enfermeras de 30-55 años que se inició en 1976 en los Estados Unidos. Desde entonces, cada dos años se han enviado a estas mujeres cuestionarios por correo. El cuestionario de 1982 incluía preguntas sobre factores de riesgo para melanoma cutáneo. En 1984 se realizó un estudio caso-control anidado, que incluyó 34 casos de melanoma cutáneo diagnosticados después de devolver el cuestionario de 1982 y 234 controles elegidos aleatoriamente de los miembros de la cohorte sin una historia de cáncer y que respondieron al mismo cuestionario. En el cuestionario del estudio caso-control se volvieron a incluir dos preguntas del cuestionario de 1982: el color del pelo y la capacidad de bronceado (Weinstock et al., 1991). Las respuestas indicadas en los dos cuestionarios se presentan en la Tabla 13.10.

	Cuestionario 1982		Cuestionario caso-control	
	Casos	Controles	Casos	Controles
Color de pelo				
Rojo o rubio	11	37	11	41
Castaño o negro ^b	23	197	23	193
	OR=2,5 (95% IC=1,1-5,7)		OR=2,3 (95% IC=1,0-5,0)	
Capacidad de bronceado				
No bronceado o claro	9	79	15	77
Medio, profundo u oscuro ^b	25	155	19	157
	OR=0,7 (95% IC=0,3-1,5)		OR=1,6 (95% IC=0,8-3,5)	

^a Adaptado de Weinstock et al. (1991).
^b Categoría de referencia.

Tabla 13.10.

Color de pelo y capacidad de bronceado autoreferidos, antes y después del diagnóstico de melanoma cutáneo en un estudio caso-control anidado^a.

exposición del individuo. Pero esto no siempre es factible (p.e. cuando se evalúan tratamientos quirúrgicos o quimioterapias con efectos secundarios visibles).

Ejemplo 13.10. Consideremos un ensayo hipotético en el que 20.000 hombres se distribuyeron aleatoriamente para recibir una intervención nueva o placebo. A continuación se siguieron durante 5 años para evaluar si la intervención podía prevenir el cáncer de pulmón. No hubo pérdidas en el seguimiento. Se produjeron un total de 75 casos de cáncer de pulmón durante el periodo de seguimiento. Sin clasificación errónea del desenlace (es decir, suponiendo una sensibilidad y especificidad del 100% en la determinación del desenlace) los resultados serían como se muestran en la Tabla 13.11.

Tabla 13.11.

Resultados de un ensayo de intervención hipotético: situación real del resultado (sensibilidad = 100% y especificidad = 100% para los grupos placebo y de intervención).

	Intervención		Total
	Nuevo	Placebo	
Casos	25	50	75
No casos	9.975	9.950	19.925
Total	10.000	10.000	20.000
Riesgo real en el grupo de intervención nuevo (r_1) = 25/10.000 = 25 por 10.000			
Riesgo real en el grupo placebo (r_0) = 50/10.000 = 50 por 10.000			
Razón de riesgo real = $r_1/r_0 = 0,50$			
Diferencia de riesgo real = $r_0 - r_1 = 25$ por 10.000			
Fracción prevenible real = $(r_0 - r_1)/r_0 = 25$ por 10.000 / 50 por 10.000 = 0,50 = 50%			

Supongamos que sólo un 80% de los casos de cáncer de pulmón se identificaron correctamente (sensibilidad=0,80) entre los que recibieron y los que no recibieron la intervención. Esto daría los resultados presentados en la Tabla 13.12.

Tabla 13.12.

Sesgo debido a la clasificación errónea no diferencial de la enfermedad: estado del desenlace observado (sensibilidad = 80% y especificidad = 100% para los grupos placebo e intervención).

	Intervención		Total
	Nuevo	Placebo	
Casos	$25 \times 0,80 = 20$	$50 \times 0,80 = 40$	60
No casos	$9.975 + (25 \times 0,20) = 9.980$	$9.950 + (50 \times 0,20) = 9.960$	19.940
Total	10.000	10.000	20.000
r_1 observada = 20 por 10.000			
r_0 observada = 40 por 10.000			
Razón de riesgos observada = $r_1/r_0 = 0,50$			
Diferencia de riesgo observada = $r_0 - r_1 = 20$ por 10.000			
Fracción prevenible observada = $(r_0 - r_1)/r_0 = 20$ por 10.000 / 40 por 10.000 = 0,50 = 50%			

Clasificación errónea no diferencial de la enfermedad

La clasificación errónea de la enfermedad es no diferencial si es similar para los individuos expuestos y para los no expuestos. Este tipo de clasificación errónea no afecta a las estimaciones del efecto relativo (si la especificidad = 100%; sensibilidad < 100%), o introduce un sesgo hacia el efecto relativo de 1 (si la especificidad < 100%; sensibilidad 100%).

Cuando la especificidad=100%, como en el Ejemplo 13.10, una falta de sensibilidad no da lugar a un sesgo en la estimación de la razón de riesgos o la fracción prevenible. No obstante, la estimación de la diferencia de riesgo está sesgada.

Ejemplo 13.11. Volvamos a considerar el ensayo hipotético presentado en la Tabla 13.11. Supongamos ahora que todos los casos de cáncer de pulmón se identificaron correctamente (sensibilidad = 100%) pero que un 10% de los sujetos sin cáncer de pulmón en cada grupo de estudio se clasificaron erróneamente como casos cáncer de pulmón (especificidad = 90%). Los resultados ahora se presentarían como en la Tabla 13.13.

	Intervención		Total
	Nuevo	Placebo	
Casos	$25 + (9.975 \times 0,10) = 1.022$	$50 + (9.950 \times 0,10) = 1.045$	2.067
No casos	$9.975 \times 0,90 = 8.978$	$9.950 \times 0,90 = 8.955$	17.933
Total	10.000	10.000	20.000

r_1 observada = 1,022 por 10.000
 r_0 observada = 1,045 por 10.000
Razón de riesgos observada = $r_1 / r_0 = 0,98$
Diferencia de riesgo observada = $r_0 - r_1 = 23$ por 10.000
Fracción prevenible observada = $(r_0 - r_1) / r_0 = 23 \text{ por } 10.000 / 1,045 \text{ per } 10.000 = 0,022 = 2,2\%$

Tabla 13.13.

Sesgo debido a la clasificación errónea no diferencial de la enfermedad: estado del desenlace observado (sensibilidad = 100% y especificidad = 90% para los grupos placebo e intervención).

En el Ejemplo 13.11, una disminución de la especificidad dio lugar a una razón de riesgos próxima a la unidad y a una subestimación considerable de la fracción prevenible.

Clasificación errónea diferencial de la enfermedad

La clasificación errónea de la enfermedad es *diferencial* si la clasificación errónea difiere entre los sujetos expuestos y los no expuestos. La clasificación errónea de enfermedad diferencial introduce un sesgo hacia una infra- o sobreestimación del efecto real de la exposición.

La clasificación errónea diferencial de la enfermedad se introduce si la exposición influencia el seguimiento e identificación de los casos. Por ejemplo, puede que los sujetos expuestos estén más (o menos) dispuestos que los sujetos no expuestos a comunicar los síntomas de la enfermedad o a visitar un médico. Asimismo, el personal implicado en el seguimiento y

diagnóstico de la enfermedad puede estar influenciado por el conocimiento del estado de exposición de los sujetos (*sesgo del observador*). Una manera de reducir al mínimo este tipo de sesgo es mantener a los observadores "ciegos" en relación a la exposición de los individuos. Pero esto no es muy factible para algunas exposiciones (p.e. intervenciones quirúrgicas, intervenciones con efectos secundarios típicos).

Ejemplo 13.12. *En Inglaterra y Gales se realizó un estudio de cohorte histórico para examinar la relación entre la exposición laboral al amianto y la mortalidad por cáncer de pulmón. Se obtuvo información sobre las causas de muerte en los certificados de defunción complementados por informes de autopsia. Se observaron un total de 11 muertes por cáncer de pulmón en los trabajadores de amianto en comparación con sólo 0,8 esperadas basándose en la mortalidad de los hombres en la población general de Inglaterra y Gales (SMR (O/E) = 14). No obstante, es probable que las autopsias de trabajadores de amianto fueran mucho más frecuentes que en la población general ya que la amiantosis es una enfermedad ocupacional por la cual las personas pueden recibir una compensación (Doll, 1955).*

En el Ejemplo 13.12, el riesgo de cáncer de pulmón en trabajadores de amianto puede haber estado sobreestimado ligeramente debido a que es probable que la causa de muerte se haya determinado más detenidamente en los trabajadores de amianto que en la población general.

13.1.3 ¿Cómo identificar el sesgo en estudios epidemiológicos?

El sesgo es una consecuencia de defectos en el diseño o ejecución de un estudio epidemiológico. El sesgo no puede controlarse en el análisis de un estudio y no puede eliminarse aumentando el tamaño de la muestra a excepción de los introducidos por la clasificación errónea no diferencial de la exposición o de la enfermedad. En capítulos anteriores se comentaron aproximaciones sobre cómo reducir el sesgo potencial en diferentes tipos de estudios epidemiológicos (Capítulos 7-11). El Cuadro 13.1 destaca algunas de las cuestiones que ayudan a identificar la presencia de sesgos en los estudios epidemiológicos.

Además de poder identificar las fuentes potenciales de sesgo en un estudio en particular, también es importante poder estimar su dirección y magnitud más probables. Para evaluar el efecto de un sesgo potencial pueden introducirse intencionadamente algunas estrategias en el estudio. Por ejemplo, en un estudio de mortalidad, el estado vital de las personas que se perdieron en el seguimiento puede determinarse mediante registros de estadísticas vitales rutinarias y su mortalidad puede compararse con la de las personas que participaron en el estudio. Una vez más, es esencial que se aplique el mismo tipo de procedimiento a todos los participantes independientemente de su estado de exposición o enfermedad.

Cuadro 13.1. Cómo evaluar la presencia de sesgos en estudios epidemiológicos

- Sesgo de selección
 - ¿Estaba la población del estudio definida claramente?
 - ¿Cuáles fueron los criterios de inclusión y exclusión?
 - ¿Se redujeron al mínimo las negativas a participar, las pérdidas del seguimiento, etc.?
- En estudios de cohorte e intervención:*
 - ¿Son los grupos similares excepto para el estado de exposición/intervención?
 - ¿Es el seguimiento adecuado? ¿Es similar para todos los grupos?
- En estudios caso-control:*
 - ¿Representan los controles a la población de donde proceden los casos?
 - ¿Estuvo la identificación y selección de los casos y controles influenciada por el estado de exposición?
- Sesgo en la medida
 - ¿Se definieron claramente las exposiciones/desenlaces de interés utilizando criterios estándar?
 - ¿Eran las mediciones lo más objetivas posible?
 - ¿Se mantuvieron los sujetos y observadores "ciegos" en relación al estado de exposición/enfermedad de los participantes y a las hipótesis del estudio?
 - ¿Estaban los observadores y los entrevistadores bien preparados?
 - ¿Se utilizaron protocolos claramente escritos para estandarizar los procedimientos en la recopilación de los datos?
 - ¿Se distribuyeron aleatoriamente los sujetos del estudio a los observadores/-entrevistadores?
 - ¿Se validó la información proporcionada por el paciente con registros o archivos existentes?
 - ¿Se utilizaron ensayos de laboratorio validados para medir la(s) exposición(es)/resultado(s)?
- ¿Se trazaron estrategias en el diseño del estudio para permitir evaluar la magnitud y dirección posible del sesgo?

13.2 ¿Podría el efecto observado ser debido a confusión?

La confusión se produce cuando una estimación de la asociación entre una exposición y un resultado se mezcla con el efecto de otra exposición con el mismo resultado, estando las dos exposiciones correlacionadas.

Por ejemplo, el fumar tabaco puede confundir las estimaciones de la asociación entre el trabajo en una profesión particular y el riesgo de cáncer de pulmón. Si se comparan las tasas de mortalidad para el cáncer de pulmón en el grupo laboral con las de la población general (de composición de sexo

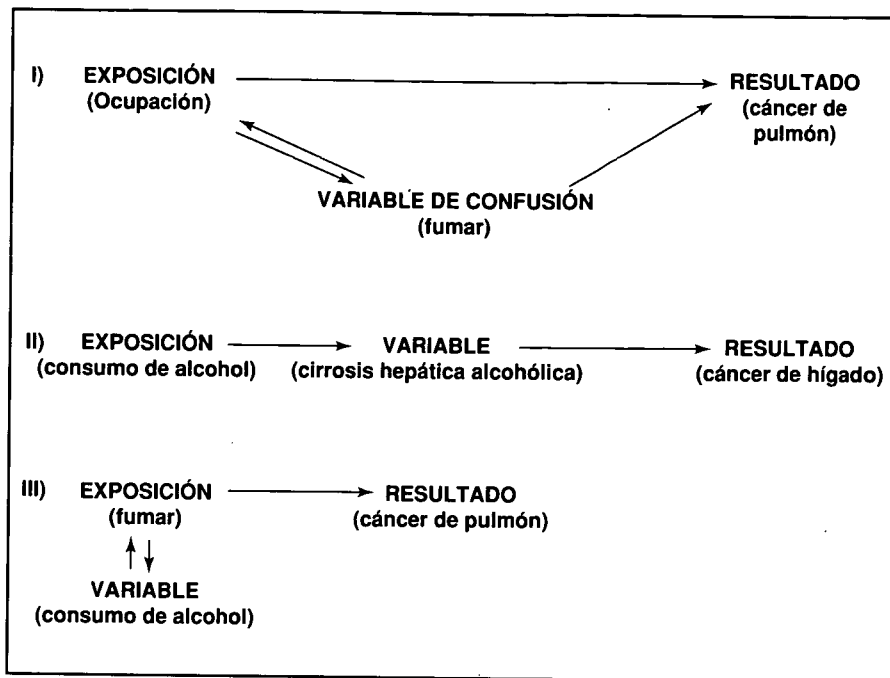


Figura 13.3. Diagrama que ilustra la relación entre exposición, resultado y variables potenciales de confusión.

y edad similar) puede observarse que el grupo laboral presenta un riesgo aumentado de cáncer de pulmón. Esto puede dar lugar a la inferencia de que la profesión es una causa directa de cáncer de pulmón. No obstante, sin más análisis, esta inferencia no sería válida si las personas empleadas en dicha profesión fumaran más que los miembros de la población general.

Para que una variable sea una *variable de confusión*, debe estar asociada con la exposición en estudio y también debe ser un factor de riesgo independiente de la enfermedad.

En la Figura 13.3, la confusión se produce sólo en el Ejemplo I

–el fumar está asociado con la ocupación concreta en estudio y es de por sí un factor de riesgo para el cáncer de pulmón. En el Ejemplo II, la cirrosis hepática alcohólica es un factor intermedio en la vía causal entre la exposición (consumo de alcohol) y la enfermedad (cáncer de hígado). En el Ejemplo III, el consumo de alcohol está asociado con la exposición en estudio (fumar) pero no es un factor de riesgo para la enfermedad (cáncer de pulmón).

Una variable de confusión potencial es todo factor que se considere pueda tener un efecto real sobre el riesgo de la enfermedad en investigación. Esto incluye tanto a los factores que tienen una relación causal directa con la enfermedad (p.e. fumar y cáncer de pulmón), como los factores que representan buenas medidas, aunque indirectas, de causas más directas desconocidas (p.e. la edad y la clase social).

La confusión puede tratarse en el diseño del estudio o en el análisis, siempre que se hayan recopilado los datos pertinentes. Pueden utilizarse tres aproximaciones para controlar la confusión en el diseño de un estudio epidemiológico:

Aleatorización – Es el método ideal para controlar las variables de confusión ya que asegura que la distribución de variables de confusión conocidas y, más importante, las desconocidas será similar en los grupos a comparar, siempre que el tamaño de la muestra sea relativamente grande. Sin embargo, este método sólo puede utilizarse en estudios experimentales.

Restricción – Es un procedimiento que limita la participación en el estudio a personas similares respecto a la variable de confusión. Por ejemplo,

si la participación en un estudio está limitada a no fumadores, se eliminará todo efecto de confusión potencial del fumar.

Apareamiento – Un procedimiento en el que los controles se seleccionan de tal modo que la distribución de las variables de confusión potenciales (p.e. edad, sexo, raza o lugar de residencia) será idéntica a la de los casos. Esto puede llevarse a cabo seleccionando para cada caso uno o más controles con características similares (p.e. de la misma edad, sexo) (apareamiento individual) o asegurando que los controles, como grupo, presenten unas características similares a los casos (apareamiento de frecuencia). En la práctica, el apareamiento se utiliza sólo en estudios caso-control (véase Apartado 9.3.4) porque sería demasiado costoso aparear sujetos en estudios de cohorte o intervención. Una excepción es el apareamiento utilizado en ensayos de intervención comunitarios (véase Apartado 7.9.2).

La confusión en los resultados puede controlarse también en el análisis utilizando:

Estratificación – Es una técnica mediante la cual la fuerza de la asociación se mide por separado dentro de cada categoría definida y homogénea (*estrato*) de la variable de confusión. Por ejemplo, si la edad es una variable de confusión, la asociación se estima por separado en cada grupo de edad; los resultados pueden agruparse utilizando un criterio adecuado para obtener una medida global resumen de la asociación, que está *ajustada o controlada* por los efectos de la variable de confusión (es decir, que tiene en cuenta las diferencias entre los grupos en la distribución de las variables de confusión; véase Capítulo 14). Cabe destacar que la estandarización, una técnica mencionada en el Capítulo 4, es un ejemplo de estratificación.

Modelos estadísticos – existen métodos estadísticos más sofisticados, como los modelos de regresión, para controlar la confusión. Son especialmente útiles cuando es necesario ajustar simultáneamente por varios factores de confusión. Estas técnicas se introducen brevemente en el Capítulo 14.

En el análisis de un estudio sólo pueden controlarse los factores de confusión si en el trabajo de campo se habían recopilado datos sobre la exposición a los factores de confusión. La magnitud en que la confusión puede controlarse dependerá de la exactitud de estos datos. Por ejemplo, la clasificación errónea no diferencial (aleatoria) de exposición a una variable de confusión subestimaré el efecto de la variable de confusión y, por consiguiente, atenuará el grado en que la confusión pueda controlarse. La asociación permanecerá incluso después del ajuste debido a la *confusión residual*. Sin embargo, a diferencia de la *clasificación errónea no diferencial de exposición o enfermedad*, la *clasificación errónea no diferencial de una variable de confusión ocasionará un sesgo en una de las dos direcciones*, dependiendo de

la dirección de la confusión. Por ejemplo, en un estudio de factores de riesgo de cáncer de cuello uterino la asociación entre el hábito de fumar y el cáncer de cuello uterino persistió incluso después de controlar en el análisis por el número de parejas sexuales. Esto puede ser debido, al menos en parte, a la confusión residual ya que el 'número de parejas sexuales' es una medida inexacta de la infección por el virus del papiloma humano, el factor de riesgo principal para el cáncer de cuello uterino.

13.3 ¿Podría el efecto observado ser debido al azar?

La evaluación del papel del azar en la interpretación de los resultados de estudios epidemiológicos se trató en el Capítulo 6. En resumen, el papel del azar puede evaluarse efectuando tests estadísticos de significación adecuados y calculando los intervalos de confianza en torno a las estimaciones del riesgo.

Un test estadístico de significación del efecto de una exposición sobre un resultado proporciona la probabilidad (P o valor de P) de que un resultado tan extremo o más que el observado se produjera sólo por azar, es decir, si no hubiera una relación real entre la exposición y el resultado. Si esta probabilidad es muy pequeña se suele declarar que el azar es una explicación poco probable de la asociación observada y por consiguiente, que existe una asociación '*estadísticamente significativa*' entre exposición y resultado. Si el valor de P es grande (normalmente mayor de 0,05), normalmente se considera que el azar no puede excluirse como explicación de la asociación observada.

Al interpretar los resultados de los tests estadísticos deben considerarse algunos aspectos. En primer lugar, el valor de P depende de la magnitud de la asociación y del tamaño del estudio. Por tanto en estudios de pequeño tamaño puede obtenerse un valor de P que no sea estadísticamente significativo, a pesar de que el efecto real de la exposición entre poblaciones sea grande. Por el contrario, con una muestra de estudio grande, efectos pequeños, que pueden ser clínica y epidemiológicamente irrelevantes, pueden alcanzar fácilmente la significación estadística. Los intervalos de confianza aportan más información que los valores de P porque proporcionan un rango de valores para la asociación exposición-resultado, que es probable incluya el efecto real sobre la población (normalmente con una probabilidad del 0,95, es decir, del 95%). También indicarán si un resultado no significativo es o no compatible con un efecto real que no se había detectado porque el tamaño de la muestra era demasiado pequeño (véase Capítulo 6). Los cálculos del tamaño de la muestra deben realizarse siempre en la fase de diseño para asegurar que el estudio tenga un *poder* suficiente (es decir, el tamaño de la muestra sea lo suficientemente grande) para detectar el efecto hipotético o la suficiente precisión para cuantificarlo (véase Capítulo 15).

En segundo lugar, los investigadores tienden a recopilar datos sobre muchos factores de riesgo potenciales y a realizar múltiples pruebas estadísticas para determinar si alguna de estas variables está asociada significativamente con la enfermedad o el resultado de interés. No obstante, la probabilidad

de obtener un valor de P estadísticamente significativo sólo por azar aumenta con el número de pruebas realizadas. Es muy importante tener esto en cuenta cuando se interpretan los datos, especialmente cuando se encuentra una asociación inesperada que no se había considerado como hipótesis al inicio del estudio. Esta asociación inesperada debería reproducirse en estudios específicamente diseñados para probarla antes de que pueda aceptarse como real.

Por último, los métodos estadísticos evalúan sólo los efectos de la variación atribuible al muestreo pero no pueden controlar los errores que no son de muestreo, tales como la confusión y el sesgo en el diseño, realización o análisis de un estudio.

13.4 ¿Podría ser causal el efecto observado?

Si el sesgo, la confusión y el azar no parecen explicar la asociación observada, ¿puede concluirse que es probable que la asociación sea causal? En un artículo publicado en 1965, Bradford Hill mencionó aspectos que requieren ser considerados al evaluar si es probable que una asociación sea causal (Hill, 1965). A continuación consideraremos cada uno de estos aspectos. Como ilustración, mostraremos cómo se aplicaron estas pautas para evaluar la asociación entre fumar y cáncer de pulmón.

Relación temporal

Éste es un criterio esencial. Para que una exposición sea la causa de una enfermedad, tiene que preceder a la aparición biológica de la enfermedad. Esto generalmente es más fácil de establecer a partir de estudios de cohorte o intervención que a partir de estudios transversales o caso-control, en los que se efectúan a la vez las medidas de exposición y enfermedad.

«El criterio de temporalidad requiere que el fumar cigarrillos sea anterior a la aparición del cáncer. Los principales estudios [de cohorte] prospectivos en los que un número enorme de sujetos inicialmente exentos de la enfermedad se siguieron durante diferentes intervalos de tiempo justifican este criterio.» (*US Surgeon General*, 1982).

Plausibilidad biológica

Es más probable que la asociación sea causal si se corresponde con otros conocimientos biológicos (p.e. experimentos en animales, mecanismos biológicos, etc.). No obstante, la ausencia de este criterio no excluye la causalidad porque la falta de plausibilidad puede reflejar simplemente la falta de conocimiento científico o el hecho de que los seres humanos son diferentes a los animales desde el punto de vista biológico.

«Se han inducido tumores benignos y malignos en la laringe de hámsters por exposición a largo plazo a humo de tabaco diluido. A pesar de los numerosos esfuerzos con varias especies y cepas, los intentos de inducir un número significativo de cánceres de pulmón en animales de labora-

torio fueron negativos. Ni ratas, ni hámsters, ni monos inhalan el humo del tabaco tan profunda e intensamente como los fumadores quienes han proporcionado la información esencial sobre las consecuencias de su 'experimento' en forma de evidencia clínica recopilada por los epidemiólogos.» (*US Surgeon General*, 1982).

Concordancia

Si se observan resultados similares en poblaciones diferentes utilizando diseños de estudio diferentes, es más probable que la asociación sea causal ya que es poco probable que todos los estudios estuvieran sujetos al mismo tipo de sesgo y/o confusión. No obstante, la falta de concordancia no excluye una asociación causal ya que en algunos estudios diferentes los niveles de exposición y otras condiciones pueden reducir el impacto del factor causal.

«Más de 50 estudios [caso-control] retrospectivos han documentado patrones de fumar (por tipo y cantidad de tabaco fumado, duración del hábito y práctica de inhalación) en series distintas de sujetos con cáncer de pulmón (p.e., hombres y mujeres, grupos ocupacionales diferentes, pacientes hospitalizados, casos de autopsia, todos los individuos que murieron de cáncer de pulmón en una zona, muestra nacional de individuos que murieron por cáncer de pulmón y razas y grupos étnicos diferentes). Muchos de estos sujetos se han comparado con controles tomados también de diferentes grupos (p.e., individuos sanos, pacientes hospitalizados por cáncer u otras enfermedades, muertos por cánceres de otros tipos y muestras de la población general). Independientemente del método, estos estudios han encontrado de manera consistente una asociación entre fumar y cáncer de pulmón. Los riesgos relativos para fumadores son siempre mayores que para no fumadores (...).

Ocho estudios [de cohorte] prospectivos importantes han examinado la relación entre fumar y mortalidad por cáncer de pulmón en un gran número de sujetos, en países diferentes y en periodos de tiempo diferentes. Los resultados de estos estudios (...) son concordantes entre sí y con los estudios [caso-control] retrospectivos.» (*US Surgeon General*, 1982).

Fuerza

La fuerza de una asociación se mide por la magnitud del riesgo relativo. Una asociación fuerte no se explica fácilmente por fuentes potenciales de sesgo o confusión y por tanto es más probable que sea causal que una asociación débil, que podría ser más fácilmente el resultado de confusión o sesgo. Por ejemplo, un riesgo relativo de magnitud 10 es más probable que refleje una asociación causal real que uno de 1,5.

«Estudios [de cohorte] prospectivos han mostrado que la tasa de mortalidad por cáncer de pulmón en fumadores es aproximadamente 10 veces la tasa de no fumadores (...). Para justificar un riesgo relativo tan alto en

términos de una asociación indirecta sería necesario que existiera un factor causal desconocido con una frecuencia como mínimo 10 veces mayor entre fumadores (...) que entre no fumadores. Un factor de confusión así debe ser fácilmente detectable y, si no se puede detectar o inferir, la observación de una asociación tan fuerte hace que sea más probable una conclusión de causalidad.» (*US Surgeon General*, 1982).

Relación exposición-respuesta

La evidencia para una relación causal es más fuerte si niveles crecientes de exposición están asociados con riesgos crecientes de enfermedad.

«La medida de exposición más fuerte encontrada por la mayoría de los estudios epidemiológicos fue para el número de cigarrillos fumados al día en el momento de la inclusión en el estudio. No obstante, otras medidas importantes de dosificación incluyen la edad en la que se empezó a fumar, la duración del hábito y la práctica de inhalación (...). Los datos (...) indican que a medida que aumenta el número de cigarrillos fumados al día existe un gradiente del riesgo de mortalidad por cáncer de pulmón. (...) Los fumadores varones que fumaban más de 20 cigarrillos al día presentaban una razones de mortalidad por cáncer de pulmón de 15 a 25 veces mayores que los no fumadores. Se observaron resultados similares entre la mujeres fumadoras (...). Cuatro estudios [de cohorte] prospectivos que examinaron la mortalidad por cáncer de pulmón según la edad a que se empezó a fumar (...) muestran una fuerte relación inversa (...), es decir, cuanto menor es la edad a que se empezó a fumar, mayor es la tasa de mortalidad por cáncer de pulmón.» (*US Surgeon General*, 1982).

Especificidad

Si una exposición particular aumenta el riesgo de una enfermedad determinada pero no el riesgo de otras enfermedades, esto es una evidencia fuerte a favor de una relación causa-efecto. No obstante, la falta de especificidad no debe utilizarse para rechazar una relación causal.

«El humo del tabaco es una mezcla compleja de varios miles de sustancias químicas. Estas sustancias diversas son capaces de reproducir más de una respuesta biológica individual. La especificidad de la asociación entre fumar y cáncer de pulmón se evidencia por la comparación de la magnitud de las razones de mortalidad por cáncer de pulmón con la de otros cánceres, como se ha hecho en la mayoría de los estudios [de cohorte] prospectivos. Las razones de mortalidad por cáncer de pulmón y tabaco son muy altas cuando se comparan con las de otros cánceres.» (*US Surgeon General*, 1982).

Reversibilidad

Cuando la eliminación de una posible causa da lugar a una reducción del riesgo de la enfermedad, aumenta la probabilidad de que la asociación sea causal. Lo ideal sería que esto se evaluara realizando un ensayo de in-

tervención aleatorizado. Lamentablemente, para muchas exposiciones/enfermedades estos ensayos aleatorios son imposibles en la práctica.

«Debido a que el hábito de fumar está asociado significativamente con el cáncer de pulmón, es lógico esperar que el dejar de fumar daría lugar a una disminución de las tasas de mortalidad por cáncer de pulmón en los que dejan de fumar en comparación con los que continúan fumando. De hecho, todos los estudios principales que examinaron la interrupción mostraron esta disminución en el riesgo de cáncer de pulmón (...). Después de 15 a 20 años, el riesgo de morir por cáncer de pulmón de los exfumadores disminuye gradualmente hasta un punto en que se aproxima más al riesgo de los no fumadores (...). La magnitud del riesgo residual que experimentan los exfumadores se determina en gran parte por la exposición acumulada al tabaco antes de dejar de fumar (es decir, cantidad total fumada por el individuo, edad al empezar a fumar y grado de inhalación) y varía con el número de años desde que se dejó de fumar, así como las razones para dejar de fumar (p.e., debido a síntomas de enfermedad).» (*US Surgeon General*, 1982).

Coherencia

La relación putativa de causa-efecto no debe discrepar seriamente con la historia natural y la biología de la enfermedad.

«El criterio final es la coherencia de la asociación entre fumar y cáncer de pulmón con hechos conocidos en la biología e historia natural del cáncer de pulmón. La coherencia de la asociación se fundamenta en los siguientes hechos: 1) Las diferencias de sexo en la mortalidad por cáncer de pulmón están bien correlacionadas con las diferencias de sexo en los patrones de fumar; 2) las diferencias de la población en la mortalidad por cáncer de pulmón están bien correlacionadas con las diferencias en el consumo de tabaco; 3) la tendencia temporal de la mortalidad por cáncer de pulmón está correlacionada con la tendencia temporal de la prevalencia del hábito de fumar.» (*US Surgeon General*, 1982).

Aunque estos aspectos con frecuencia se denominan 'criterios' (como en el informe del *US Surgeon General* (1982)), no se deberían considerar condiciones necesarias para establecer la causalidad. La única excepción es la temporalidad – para que una exposición sea una causa de una enfermedad debe preceder claramente a la aparición biológica de la misma.

13.5 Comentarios generales

En epidemiología, es raro que un único estudio proporcione las suficientes 'pruebas' de que una exposición determinada afecte el riesgo de una enfermedad particular. No obstante, nuestro grado de aceptación de la asociación dependerá del diseño del estudio. Los estudios ecológicos *per se* pueden mostrar asociaciones, pero debido al gran potencial de confusión nunca pueden

utilizarse para establecer la causalidad a nivel individual. Los ensayos aleatorios bien realizados son nuestras mejores herramientas para evaluar la causalidad, pero sus observaciones siempre deben interpretarse en el contexto de todas las demás evidencias disponibles, incluyendo las evidencias de otras áreas de investigación. Más aun, para algunas relaciones exposición/enfermedad, estas intervenciones aleatorizadas no son posibles o no son éticas.

Probabilidad de:	Ecológico	Transversal	Caso-control	Cohorte	Ensayo aleatorizado
Sesgo de selección					
Selección de sujetos	N/A	media	alta	baja	baja
Pérdidas en el seguimiento	N/A	N/A	baja	alta	media
Sesgo de memoria					
	N/A	alta	alta	baja	baja
Confusión					
	alta	media	media	baja	muy baja

^a Modificado a partir de Beaglehole *et al.* (1993)
N/A = No procede.

Tabla 13.14. Probabilidad de sesgo de selección, sesgo de memoria y confusión para los diferentes diseños de estudio^a.

La Tabla 13.14 proporciona una idea de cuánto pueden estar afectados por el sesgo y la confusión los resultados de diferentes tipos de estudios.

Los datos de diferentes estudios que han tratado la misma cuestión pueden reunirse y *volverlos a analizar* utilizando técnicas estadísticas especiales que consideran el hecho de que los estudios originales pueden diferir en gran cantidad de aspectos (p.e. características de los sujetos del estudio, métodos de obtención de datos). Esta estrategia es especialmente útil cuando es probable que el efecto de la exposición sea pequeño, siendo necesario un tamaño de la muestra muy grande para que se detecte y cuantifique con precisión.

Ejemplo 13.13. *En las últimas décadas se realizaron numerosos estudios de cohorte y caso-control para examinar la relación entre el uso de anticonceptivos orales y cáncer de mama. Los resultados de estos estudios fueron inconsistentes. Para clarificar este aspecto, los datos de la mayoría de estos estudios se recopilaron y volvieron a analizar. Los 54 estudios incluidos en este reanálisis se realizaron en 25 países de Europa, América del norte, Asia, Australasia, África y América latina. Juntos incluían 53.297 mujeres con cáncer de mama invasivo y 100.239 sin cáncer de mama. La Figura 13.4 muestra los riesgos relativos de cáncer de mama en las que habían usado anticonceptivos orales combinados (el tipo más utilizado) alguna vez frente a las que nunca los habían utilizado. Las estimaciones del riesgo variaron considerablemente de estudio a estudio pero la mayoría de los intervalos de confianza fueron amplios e incluyeron el valor de la hipótesis nula de 1. La estimación global era ligeramente superior a 1 (1,07) y no incluía el valor de la hipótesis nula. Las mujeres que tomaban anticonceptivos orales combinados en ese momento o aquellas que los habían dejado hacía menos de 10 años presentaban un pequeño aumento en el riesgo de cáncer de mama (Figura 13.5d) (Collaborative Group on Hormonal Factors in Breast Cancer, 1996).*

El *meta-análisis* es otra técnica utilizada para combinar los resultados de diferentes estudios. Difiere de un *reanálisis* en que no se accede a los datos originales de cada estudio individual. Sólo se dispone y se pueden utilizar las estimaciones del efecto de la exposición publicadas de cada estudio para generar un resultado global integrado.

Es esencial asegurar que prácticamente todos los estudios (publicados y no publicados) se incluyan en cualquier reanálisis o meta-análisis. Esto es debido a que los estudios publicados tienden a ser una muestra sesgada, con una representación mayor de los que mostraban resultados «positivos».

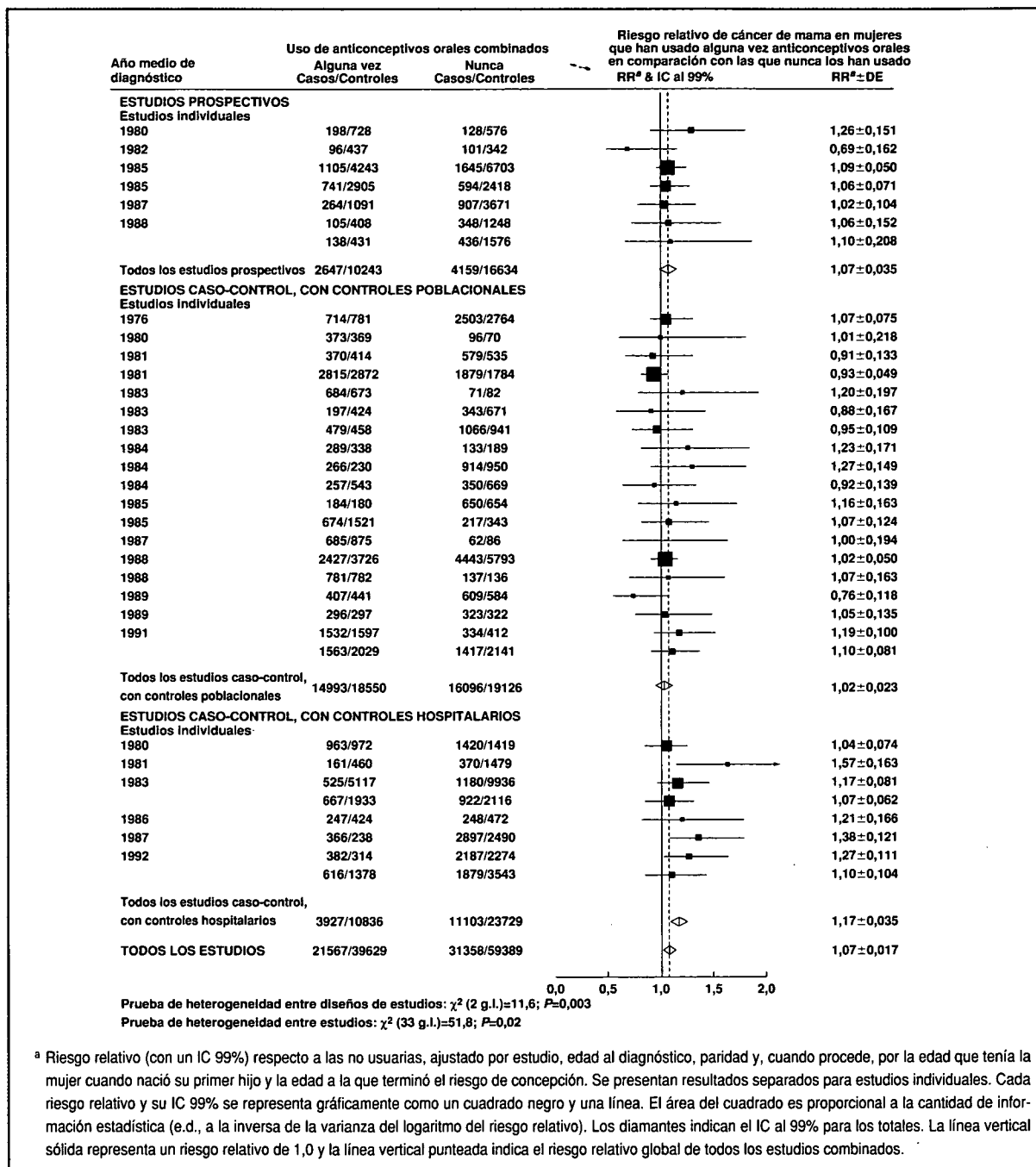


Figura 13.4. Riesgo relativo de cáncer de mama en mujeres que habían usado alguna vez anticonceptivos orales combinados en comparación con las que nunca los habían usado (reproducido con la autorización del Collaborative Group on Hormonal Factors in Breast Cancer, 1996. © The Lancet Ltd., 1996; en este artículo se incluye una discusión completa de los métodos y una referencia completa a los estudios individuales).

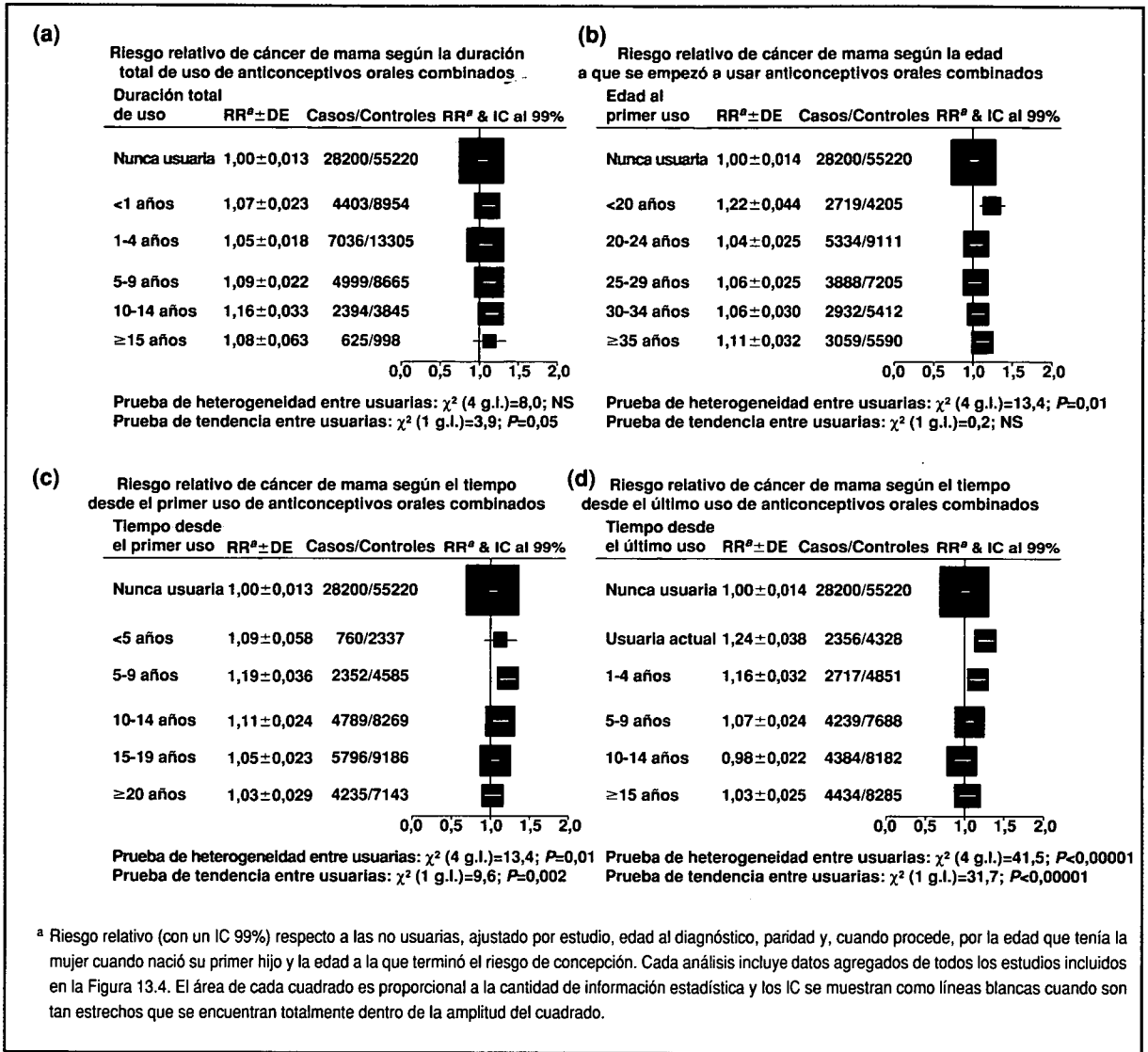


Figura 13.5.

Riesgo relativo de cáncer de mama para varios índices de tiempo de uso de anticonceptivos orales combinados (reproducido con la autorización del Collaborative Group on Hormonal Factors in Breast Cancer, 1996, © por The Lancet Ltd., 1996; en este artículo se incluye una discusión completa de los métodos con la referencia completa de cada estudio individual).

Cuadro 13.2. Puntos esenciales

- En la interpretación de los resultados de cualquier estudio epidemiológico es necesario considerar el sesgo, la confusión y el azar.
- El sesgo tiende a dar lugar a una estimación incorrecta del efecto de una exposición sobre el resultado de interés. Existen dos tipos principales de sesgo:
 - El *sesgo de selección* se produce cuando existe una diferencia entre las características de las personas que participaron en el estudio y las características de aquellas que no participaron. El sesgo de selección puede ser un problema importante en estudios caso-control aunque también puede afectar a estudios transversales y, en menor grado, a estudios de cohorte y ensayos aleatorios.
 - El *sesgo en la medida* se produce cuando las determinaciones o clasificaciones de la enfermedad o exposición no son válidas. Las consecuencias del sesgo de medida dependen del carácter diferencial o no diferencial del error en la medida y de si afecta a la exposición o al resultado de interés.
- La confusión se produce cuando una estimación de la asociación entre una exposición y un resultado se mezclan con el efecto real de otra exposición en el mismo resultado, estando las dos exposiciones correlacionadas. Existen varias maneras de tratar la confusión en la fase de diseño (aleatorización, restricción, apareamiento) y en el análisis (estratificación y modelo estadístico).
- El papel del azar se evalúa mediante un test estadístico y, sobretodo, calculando los intervalos de confianza.
- Incluso si el sesgo, confusión y azar no parecen explicar la asociación observada, debe evaluarse la probabilidad de que una asociación sea causal aplicando los criterios propuestos por Hill (1965).

Lecturas adicionales

* Los aspectos a considerar cuando se evalúa una relación exposición-resultado los presentó Bradford Hill en un artículo publicado en 1965.

Capítulo 14

Cómo tratar la confusión en el análisis

En el capítulo anterior se comentó brevemente cómo podía tratarse la confusión tanto en la fase de diseño de un estudio como durante el análisis de los resultados. Más adelante se mencionó que existen dos procedimientos estadísticos principales que pueden utilizarse en el análisis – *estratificación y modelos mediante regresión*. En este capítulo comentaremos estas dos técnicas con más detalles. Obviamente, estos métodos sólo pueden aplicarse si se dispone de datos sobre los factores potenciales de confusión. Por tanto, las variables potenciales de confusión tienen que identificarse pronto en la fase de diseño del estudio para asegurar que se obtiene una información válida sobre las mismas.

14.1 Estratificación

Un factor de confusión es aquél que está relacionado simultáneamente con la variable de exposición y la variable de resultado y que no se encuentra en la vía causal entre ellas (Apartado 13.2). Ignorar la confusión al evaluar la asociación entre una variable de exposición y una variable de resultado puede dar lugar a una sobrestimación o infraestimación de la asociación real entre la exposición y el resultado e incluso puede cambiar la dirección del efecto observado.

Ejemplo 14.1. Imaginemos un estudio caso-control hipotético para examinar la relación entre fumar y cáncer de ovario entre mujeres nulíparas. La Tabla 14.1 muestra los resultados de este estudio.

	Fumar		Total
	Si	No	
Casos de cáncer de ovario	24 (a)	36 (b)	60 (n ₁)
Controles	58 (c)	40 (d)	98 (n ₀)
Total	82 (m₁)	76 (m₀)	158 (N)

Odds ratio cruda = $(24/36) / (58/40) = 0,46$
Intervalo de confianza al 95% = 0,23–0,93
Test $\chi^2 = 5,45$ con 1 g.l.; $P = 0,02$

Tabla 14.1. Resultados de un estudio caso-control sobre fumar y cáncer de ovario: datos hipotéticos.

En el Ejemplo 14.1, las mujeres con cáncer de ovario presentan una prevalencia de fumar mucho menor ($24/60 = 40\%$) en comparación con los controles ($58/98 = 59\%$). Esto sugiere que el fumar protege frente al cáncer

de ovario (*odds ratio* (OR) = 0,46). Como se ha comentado en el capítulo anterior, existen varias explicaciones posibles para esta observación:

(i) *Sesgo*: la *odds ratio* observada de 0,46 no representa exactamente la *odds ratio* real debido a un sesgo de selección o de medida.

(ii) *Azar*: la asociación observada entre fumar y cáncer de ovario ha aparecido por azar. El intervalo de confianza al 95% alrededor de la *odds ratio* observada es igual a 0,23 – 0,93 y el test χ^2 proporciona una $P = 0,02$. Por tanto, el azar es una explicación poco probable para la observación.

(iii) *Confusión*: la *odds ratio* observada de 0,46 es debida al efecto de otra variable. Por ejemplo, puede ser que las mujeres que fumaban fueran diferentes en otros aspectos de las no fumadoras y presentaran una probabilidad menor de desarrollar cáncer de ovario debido, más que al tabaco, a estos otros aspectos no conocidos.

(iv) *Causa*: fumar reduce el riesgo de cáncer de ovario. El intervalo de confianza al 95% indica la precisión de la estimación del riesgo obtenida a partir de la muestra en relación al efecto real en la población.

En el Ejemplo 14.1, es posible que aparezca una asociación entre fumar y cáncer de ovario debido al efecto de confusión de otros factores tales como el uso de anticonceptivos orales. Los resultados mostrados en la Tabla 14.1 incluyen a todas las mujeres combinadas independientemente de su historia de uso de anticonceptivos orales. Para evaluar si el uso de anticonceptivos orales es un factor de confusión, debemos observar la asociación entre fumar y cáncer de ovario por separado para las mujeres que usan anticonceptivos orales y las que nunca los han usado. Este análisis se muestra en la Tabla 14.2.

Tabla 14.2.
Estudio caso-control hipotético sobre fumar y cáncer de ovario descrito en el Ejemplo 14.1: resultados presentados por separado para las mujeres que nunca han usado y las que han usado alguna vez anticonceptivos orales.

Nunca usuarias de anticonceptivos orales			
	Fumar		Total
	Sí	No	
Casos de cáncer de ovario	9 (a_1)	32 (b_1)	41 (n_{11})
Controles	8 (c_1)	28 (d_1)	36 (n_{01})
Total	17 (m_{11})	60 (m_{01})	77 (N_1)
<i>Odds ratio</i> = 0,98 (intervalo de confianza al 95% = 0,30–3,37)			
Alguna vez usuarias de anticonceptivos orales			
	Fumar		Total
	Sí	No	
Casos de cáncer de ovario	15 (a_2)	4 (b_2)	19 (n_{12})
Controles	50 (c_2)	12 (d_2)	62 (n_{02})
Total	65 (m_{12})	16 (m_{02})	81 (N_2)
<i>Odds ratio</i> = 0,90 (intervalo de confianza al 95% = 0,23–4,40)			

En cada categoría (o *estrato*) de uso de anticonceptivos orales, la prevalencia de fumar fue similar en mujeres con y sin cáncer de ovario (22% y 22% entre las nunca usuarias y 79% y 81% entre las usuarias). No obstante, cuando se combinaron las usuarias de anticonceptivos orales y las que

nunca los habían usado (Tabla 14.1), se observó una clara diferencia en la prevalencia de fumar entre casos y controles (40% frente al 59%). Dos factores eran responsables de esta observación:

1. Entre los controles, las fumadoras presentaban una prevalencia mucho mayor de uso de anticonceptivos orales (50/58 = 86%) que las no fumadoras (12/40 = 30%), es decir, había una asociación entre estas dos variables (Tabla 14.3).

		Uso de anticonceptivos orales		Total
		Alguna vez	Nunca	
Fumar	Sí	50	8	58
	No	12	28	40
	Total	62	36	98

Tabla 14.3.

Estudio caso-control hipotético descrito en el Ejemplo 14.1: distribución de los controles por hábitos de fumar y uso de anticonceptivos orales.

Obsérvese que la asociación entre fumar y uso de anticonceptivos orales se examinó en los controles en lugar de los casos, o en los dos grupos combinados. Esto es debido a que los controles deben representar a la población de donde se seleccionaron los casos y necesitamos evaluar esta asociación en la población general. En un estudio de cohorte o de intervención, la asociación se analizaría construyendo una tabla similar, sustituyendo el número de controles por personas-años a riesgo si la medida adecuada del efecto fuera una razón de tasas o el número de personas a riesgo al inicio del seguimiento si la medida del efecto fuera una razón de riesgos.

2. El uso de anticonceptivos orales es considerablemente menor en los casos de cáncer de ovario que en los controles. Los datos de la Tabla 14.2 pueden reagruparse de modo que se ignore el fumar (Tabla 14.4).

	Uso de anticonceptivos orales		Total
	Alguna vez	Nunca	
Casos de cáncer de ovario	19	41	60
Controles	62	36	98
Total	81	77	158

Tabla 14.4.

Estudio caso-control hipotético descrito en el Ejemplo 14.1: distribución de casos y controles según el uso de anticonceptivos orales.

Sólo un 32% (= 19/60) de las mujeres con cáncer de ovario eran usuarias de anticonceptivos orales mientras que un 63% (= 62/98) de los controles eran usuarias.

Debido a que en estos datos el uso de anticonceptivos orales estaba asociado tanto a la exposición (fumar) como a la enfermedad de interés (cáncer de ovario), la variable uso de anticonceptivos orales actúa como factor de confusión. Como consecuencia, al combinar los datos para usuarias y no-usuarias, el resultado indicaba una asociación entre fumar y cáncer de ovario mucho más fuerte de la que existía realmente (*confusión positiva*). En otras situaciones (como en el Ejemplo 14.2; véase apartado siguiente), la combinación de estratos en presencia de un factor de confusión puede enmascarar un efecto que existe realmente (*confusión negativa*), o incluso mostrar un efecto en la dirección contraria a la real.

14.2 Medidas resumen de efecto de Mantel-Haenszel

Cuando analizamos los resultados de un estudio para buscar evidencia de una determinada asociación entre exposición y enfermedad, normalmente empezamos incluyendo en el análisis todos los sujetos en la muestra del estudio. Este análisis proporciona una estimación *cruda* del efecto de la exposición sobre la enfermedad de interés. El siguiente paso lógico es dividir nuestra muestra de estudio en varios subgrupos o *estratos*, definidos por una variable potencial de confusión, y ver si los resultados son coherentes en cada uno de los estratos. Esta aproximación es muy informativa ya que describe cómo varía el efecto de la exposición sobre la enfermedad de interés en los subgrupos de sujetos con características diferentes. Pueden presentarse simplemente las estimaciones del efecto específico en cada estrato y sus intervalos de confianza. Se supone que cada una de estas estimaciones específicas del estrato son homogéneas respecto a la variable potencial de confusión y por tanto no están confundidas.

No obstante, normalmente no estamos muy interesados en los resultados específicos de cada estrato *per se* y preferiríamos tener una estimación global única. En otras palabras, querríamos poder calcular una estimación resumen del efecto que, a diferencia de la estimación *cruda*, considerara el efecto de confusión de la variable de estratificación. Estas estimaciones *ajustadas* pueden calcularse *combinando* los resultados de cada estrato. Pero incluso aunque el efecto real sea el mismo en todos los estratos, sería de esperar que nuestras estimaciones fueran diferentes debido a la variación aleatoria. El análisis combinado tiene en cuenta esto, dando un mayor peso a las estimaciones de efecto de los estratos más grandes. Esto implica el cálculo de una *media ponderada* de las estimaciones específicas de cada estrato eligiendo una serie de *pesos* que aumenta al máximo la precisión estadística del estimador de efecto ajustado. Existen varios procedimientos alternativos de ponderación que consiguen estimaciones del efecto resumidas y precisas. En este apartado, nos concentraremos en un procedimiento diseñado por Mantel y Haenszel que es ampliamente utilizado y relativamente simple de aplicar.

14.2.1 La *odds ratio* de Mantel-Haenszel

Volvamos a considerar el Ejemplo 14.1. Debido a que el uso de anticonceptivos orales es un factor de confusión de la relación entre fumar y cáncer de ovario, no debemos combinar para el análisis los datos de las mujeres usuarias y las no-usuarias. Por ello, la *odds ratio* cruda de 0,46 obtenida en la Tabla 14.1 no es apropiada. Podríamos calcular la *odds ratio* y sus intervalos de confianza al 95% separadas para cada grupo de uso de anticonceptivos orales tal y como se muestra en la Tabla 14.2. Pero desearíamos poder resumir los resultados globales del estudio de un modo que se elimine el efecto de confusión del uso de anticonceptivos orales. La *odds ratio* de Mantel-Haenszel, indicada por OR_{MH} , proporciona una *media ponderada de las odds ratios en los diferente estratos*, en la que se asigna más peso a las OR de los estratos más grandes.

Para calcular la OR_{MH} , empezamos construyendo tablas 2×2 de exposición y enfermedad para los diferentes estratos de la variable de confusión tal y como se muestra en la Tabla 14.2. La OR_{MH} puede obtenerse entonces aplicando la fórmula siguiente:

$$OR_{MH} = \frac{a_1 d_1 / N_1 + a_2 d_2 / N_2}{b_1 c_1 / N_1 + b_2 c_2 / N_2}$$

Si aplicamos esta fórmula a los datos de nuestro estudio caso-control de cáncer de ovario (Tabla 14.2) podemos obtener:

$$OR_{MH} = \frac{(9 \times 28) / 77 + (15 \times 12) / 81}{(32 \times 8) / 77 + (4 \times 50) / 81} = \frac{5,49}{5,79} = 0,95$$

Así, la *odds ratio* para fumar ajustada por el uso de anticonceptivos orales es igual a 0,95. Esta *odds ratio* ajustada contrasta con la *odds ratio* cruda de 0,46 obtenida a partir de la Tabla 14.1. El ajuste por uso de anticonceptivos orales proporciona una *odds ratio* mucho más próxima a la unidad, lo que significa que la protección ejercida por fumar, si existe, es mucho menos fuerte que la que nos hizo pensar el resultado inicial.

La fórmula anterior puede ampliarse fácilmente a más de dos estratos en los que ambos, numerador y denominador, se suman para todos los estratos:

$$OR_{MH} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i}$$

En esta fórmula, Σ significa suma y el subíndice i sustituye a los subíndices 1, 2, 3, etc., que representan cada uno de los estratos.

Podemos calcular un intervalo de confianza alrededor de la OR_{MH} y un test de χ^2 de Mantel-Haenszel utilizando las fórmulas indicadas en el Apartado A14.1.1. En nuestro ejemplo de cáncer de ovario, el intervalo de confianza al 95% es 0,42-2,16. El test χ^2 es igual a 0,016 con un grado de libertad, que corresponde a una $P = 0,93$. Por tanto, tras ajustar por el uso de anticonceptivos orales, el efecto de fumar sobre el cáncer de ovario deja de ser 'estadísticamente significativo'.

14.2.2 Razón de riesgos de Mantel-Haenszel

El método de Mantel-Haenszel también puede utilizarse para obtener una *razón de riesgos ajustada*. Consideremos el ejemplo siguiente.

En el Ejemplo 14.2, el análisis crudo muestra que el riesgo de desarrollar cáncer de pulmón en los expuestos a la sustancia química es un 52% mayor que en los no expuestos (Tabla 14.5). Antes de concluir que la sustancia química está asociada a un riesgo aumentado de cáncer de pulmón, es necesario excluir la posibilidad de que fumar, en lugar de la exposición ocupacional, sea la explicación de la asociación observada. Para hacer esto, es necesario examinar los datos por separado para fumadores y no fumadores.

Ejemplo 14.2. Supongamos que se realiza un estudio de cohorte para investigar si la exposición ocupacional a una sustancia química en particular estaba asociada con un riesgo aumentado de cáncer de pulmón. Todos los sujetos del estudio se siguieron durante cinco años después de la inclusión en el estudio (o hasta el diagnóstico de cáncer de pulmón, si era antes). Los resultados de este estudio se muestran en la Tabla 14.5.

Tabla 14.5. Resultados de un estudio de cohorte sobre la exposición ocupacional a una sustancia química en particular y cáncer de pulmón: datos hipotéticos.

		Exposición a una sustancia química		Total
		Sí	No	
Cáncer de pulmón	Sí	480 (a)	360 (b)	840 (n ₁)
	No	83.520 (c)	95.640 (d)	179.160 (n ₀)
	Total	84.000 (m₁)	96.000 (m₀)	180.000 (N)

Razón de riesgo cruda = (480/84.000)/(360/96.000) = 5,71 por 1.000/3,75 por 1.000 = 1,52
 Intervalo de confianza al 95% = 1,33–1,75
 $\chi^2 = 37,21$ con 1 g.l.; $P < 0,0001$

La Tabla 14.6 muestra que las razones de riesgos específicas para cada estrato son mayores que la razón de riesgos cruda (2,0 frente a 1,52). Este es un ejemplo de *confusión negativa*. Esta confusión se produjo porque la pre-

Tabla 14.6. Estudio de cohorte hipotético sobre la exposición ocupacional a una sustancia química particular y cáncer de pulmón descrito en el Ejemplo 14.2: resultados presentados por separado para fumadores y no fumadores.

Fumadores				
		Exposición a una sustancia química		Total
		Sí	No	
Cáncer de pulmón	Sí	80 (a ₁)	160 (b ₁)	240 (n ₁₁)
	No	3.920 (c ₁)	15.840 (d ₁)	19.760 (n ₀₁)
	Total	4.000 (m₁₁)	16.000 (m₀₁)	20.000 (N₁)

Razón de riesgos = (80/4.000)/(160/16.000) = 20 por 1.000 / 10 por 1.000 = 2,0
 Intervalo de confianza 95% = 1,53–2,61

No fumadores				
		Exposición a una sustancia química		Total
		Sí	No	
Cáncer de pulmón	Sí	400 (a ₂)	200 (b ₂)	600 (n ₁₂)
	No	79.600 (c ₂)	79.800 (d ₂)	159.400 (n ₀₂)
	Total	80.000 (m₁₂)	80.000 (m₀₂)	160.000 (N₂)

Razón de riesgos = (400/80.000)/(200/80.000) = 5,0 por 1.000 / 2,5 por 1.000 = 2,0
 Intervalo de confianza al 95% = 1,69–2,37

valencia de fumar, un factor de riesgo importante para el cáncer de pulmón, era mucho menor en los trabajadores expuestos a la sustancia química (4.000 / 84.000 = 5%) que en los no expuestos (16.000 / 96.000 = 17%) (Tabla 14.7).

	Exposición a una sustancia química		Total
	Sí	No	
Fumar	Sí	4.000	16.000
	No	80.000	80.000
	Total	84.000	96.000

Tabla 14.7. Estudio de cohorte hipotético descrito en el Ejemplo 14.2: distribución de los sujetos del estudio por exposición ocupacional y hábito de fumar.

Puede obtenerse una estimación resumen Mantel-Haenszel de la razón de riesgos común (RR_{MH}) aplicando la fórmula siguiente a nuestros datos

$$RR_{MH} = \frac{\sum a_i m_{0i} / N_i}{\sum b_i m_{1i} / N_i}$$

Así,

$$RR_{MH} = \frac{(80 \times 16.000) / 20.000 + (400 \times 80.000) / 160.000}{(160 \times 4.000) / 20.000 + (200 \times 80.000) / 160.000} = \frac{264}{132} = 2,0$$

Los cálculos de los intervalos de confianza alrededor de RR_{MH} y del test χ^2 de Mantel-Haenszel se muestran en el Apartado A14.1.2. En nuestro ejemplo, el intervalo de confianza al 95% es de 1,73 a 2,30. El test χ^2 es igual a 92,99 con un grado de libertad, que proporciona una $P < 0,0001$. Así, existe una fuerte evidencia de que la exposición ocupacional estaba asociada con un riesgo aumentado de cáncer de pulmón y este efecto era incluso más fuerte cuando se consideraban las diferencias en el hábito de fumar entre los trabajadores expuestos y los no expuestos.

14.2.3 Razón de tasas de Mantel-Haenszel

El método de Mantel-Haenszel también puede aplicarse cuando la medida apropiada del efecto es la razón de tasas en lugar de la razón de riesgos. Este método proporciona una razón de tasas ajustada (indicada por RT_{MH}) calculando una media ponderada de la razón de tasas en los diferentes estratos.

En el Ejemplo 14.3, la razón de tasas cruda y su intervalo de confianza son coherentes con una disminución de la incidencia de cáncer de cuello uterino en mujeres que en el cuestionario inicial indicaron haberse realizado alguna vez un frotis de Papanicolaou. Otros estudios han mostrado que existe un gradiente socioeconómico en la incidencia de cáncer de cuello uterino presentando un mayor riesgo las mujeres con un nivel socioeconómico bajo. Por tanto, los factores socioeconómicos pueden haber

Ejemplo 14.3. Supongamos que una cohorte de mujeres sanas de 45-50 años se siguió para estudiar los factores de riesgo para varios cánceres femeninos. En el momento de la inclusión en el estudio, se administró un cuestionario a las mujeres sobre variables sociodemográficas y ginecológicas y sobre su historia reproductiva. Durante el periodo de seguimiento aparecieron un total de 1.141 casos de cáncer de cuello uterino. La relación entre cáncer de cuello uterino y haberse realizado alguna vez un frotis de Papanicolaou (de acuerdo a la información facilitada en el cuestionario inicial) se muestra en la Tabla 14.8.

Tabla 14.8. Resultados de un estudio de cohorte sobre frotis de Papanicolaou y cáncer de cuello uterino: datos hipotéticos.

	Frotis de Papanicolaou		Total
	Alguna vez	Nunca	
Casos	17 (a)	1.124 (b)	1.141 (n)
Personas-años a riesgo	71.184 (y ₁)	1.518.701 (y ₀)	1.589.885 (y)
Tasa por 100.000 p-as	23,9 (r ₁)	74,0 (r ₀)	71,8 (r)
Razón de tasas cruda = (17/71.184)/(1.124/1.518.701) = 0,32			
Intervalo de confianza 95% = 0,20 – 0,52			
$\chi^2 = 23,69$ con 1 g.l.; $P < 0,001$			

confundido la asociación entre el frotis de Papanicolaou y cáncer de cuello uterino si, por ejemplo, las mujeres con un alto nivel socioeconómico presentaban una probabilidad mayor de visitar a sus médicos y de realizarse un frotis de Papanicolaou como parte de su examen médico regular. Para aclarar este aspecto, es necesario examinar la relación entre el frotis de Papanicolaou y cáncer de cuello uterino por separado para las mujeres con niveles socioeconómicos diferentes. Esto se muestra en la Tabla 14.9, en la que se utiliza el nivel educativo de la mujer como marcador del nivel socioeconómico.

Tabla 14.9. Estudio de cohorte hipotético sobre frotis de Papanicolaou y cáncer de cuello uterino descrito en el Ejemplo 14.3: resultados estratificados por el nivel educativo de la mujer.

Nivel educativo alto			
	Frotis de Papanicolaou		Total
	Alguna vez	Nunca	
Casos	13 (a ₁)	697 (b ₁)	710 (n ₁)
Personas-años a riesgo	38.346 (y ₁₁)	828.149 (y ₀₁)	866.495 (y ₁)
Tasa por 100.000 p-as	33,9 (r ₁₁)	84,2 (r ₀₁)	81,9 (r ₁)
Razón de tasas = 0,40; intervalo de confianza 95% = 0,23–0,69			
Nivel educativo bajo			
	Frotis de Papanicolaou		Total
	Alguna vez	Nunca	
Casos	4 (a ₂)	427 (b ₂)	431 (n ₂)
Personas-años a riesgo	32.838 (y ₁₂)	690.552 (y ₀₂)	723.390 (y ₂)
Tasa por 100.000 p-as	12,2 (r ₁₂)	61,8 (r ₀₂)	59,6 (r ₂)
Razón de tasas = 0,20; intervalo de confianza 95% = 0,08–0,54			

La fórmula para la estimación resumen de la razón de tasas común de Mantel-Haenszel es

$$RT_{MH} = \frac{\sum a_i \gamma_{0i} / \gamma_i}{\sum b_i \gamma_{1i} / \gamma_i}$$

Por tanto, en nuestro ejemplo

$$\begin{aligned} RT_{MH} &= \frac{(13 \times 828.149) / 866.495 + (4 \times 690.552) / 723.390}{(697 \times 38.346) / 866.495 + (427 \times 32.838) / 723.390} \\ &= \frac{16,24}{50,23} = 0,32 \end{aligned}$$

Por tanto, en estos datos, el nivel educativo no era una variable de confusión del efecto del antecedente de frotis de Papanicolaou sobre el cáncer de cuello uterino ya que las razones de tasas cruda y ajustada tienen exactamente el mismo valor (0,32).

Aplicando las fórmulas indicadas en el Anexo (Apartado A14.1.3), obtenemos un test χ^2 de Mantel-Haenszel de 23,73 con un grado de libertad, que corresponde a una $P < 0,001$. El intervalo de confianza al 95% es de 0,20 a 0,52. Por tanto, el azar es una explicación poco probable para esta observación.

Cabe destacar que el método de Mantel-Haenszel para controlar la confusión es similar al método de estandarización utilizado para calcular las tasas ajustadas por edad en el Capítulo 4. Estos dos métodos son conocidos como *análisis estratificados* porque observamos una exposición y una enfermedad de interés para cada uno de los diferentes *estratos* (niveles) de una variable de confusión. No obstante, difieren en los pesos utilizados para calcular la media ponderada de la razón de tasas en los diferentes estratos (véase Apartado 4.3.3).

14.3 Cómo identificar variables de confusión

Con objeto de poder examinar el efecto de las variables potenciales de confusión en el análisis es necesario identificarlas en la fase de diseño del estudio. La edad y el sexo son variables de confusión potenciales obvias en prácticamente todos los estudios. Fumar también será un factor de confusión potencial en todo estudio que examine la relación entre una exposición concreta y el cáncer de pulmón. En este caso es necesario excluir la posibilidad de que sea el fumar, en lugar de la exposición en estudio, el factor de riesgo responsable de cualquier asociación encontrada. Las variables potenciales de confusión también deben incluir factores tales como el nivel socioeconómico o el lugar de residencia que son medidas aproximadas de causas, más directas pero desconocidas, de enfermedad.

No todos los factores que se sospecha son factores potenciales de confusión darán lugar realmente a confusión de la relación exposición-enfermedad en un estudio en particular. Pero, ¿cómo sabemos realmente si una va-

riable en particular es un factor de confusión? Definimos un factor de confusión como un factor que está asociado a la vez con la exposición y con la enfermedad (y no está en la vía causal). Esto puede ser difícil de evaluar en la práctica. Por ejemplo, un tamaño de muestra grande podría dar lugar fácilmente a una asociación pequeña, pero estadísticamente significativa, entre el factor de confusión y la exposición o la enfermedad; no obstante, pueden no ser lo suficientemente grandes como para ocasionar un efecto de confusión. Por tanto, la presencia o ausencia de confusión no debe evaluarse utilizando un test estadístico de significación. La magnitud de la confusión en cualquier estudio debe evaluarse observando el grado de discrepancia entre las estimaciones cruda y ajustada. Si no existe una diferencia entre estas estimaciones, el efecto exposición-enfermedad observado no estaba confundido por la variable potencial de confusión. Una diferencia importante, como la observada en el Ejemplo 14.1, indica la presencia de confusión e implica que la medida resumen *ajustada* es una estimación mejor del efecto de la exposición sobre la enfermedad de interés que la medida resumen *cruda*, ya que el ajuste elimina el efecto del factor de confusión.

También deben considerarse dos aspectos de la estratificación. En primer lugar, los factores que se encuentran en la vía causal entre una exposición y una enfermedad no deben considerarse como variables de confusión de la asociación entre la exposición y el resultado. Controlar por un factor que se encuentra en la vía causal da lugar a una infraestimación de la fuerza del efecto de la exposición sobre la enfermedad en estudio. Ocasionalmente, una variable que se encuentra en la vía causal puede estar ajustada en el análisis si queremos evaluar si el efecto de la exposición sobre el resultado en un estudio está mediado completamente por esta variable intermedia o si pueden haber otros mecanismos biológicos independientes. Por ejemplo, si creemos que la infección por virus del papiloma humano (VPH) se encuentra en la vía causal entre el número de parejas sexuales y el cáncer de cuello uterino, la asociación entre ambas variables debe desaparecer tras ajustar por la infección por VPH. Si el efecto no desaparece completamente, puede indicar que el efecto del número de parejas sexuales sobre el cáncer de cuello uterino puede estar mediado por otros mecanismos biológicos no relacionados directamente con la infección por VPH. En la práctica, debido a errores en la medición de la variable intermedia, este razonamiento puede no ser tan sencillo.

En segundo lugar, es importante destacar que la estratificación supone que cada estrato es homogéneo respecto a la variable de confusión. Esto depende en parte de la validez de los datos sobre la variable de confusión (véase Apartado 13.2) y de la manera como se han definido los estratos. Por ejemplo, el control de la edad mediante estratificación se conseguirá mejor si los estratos son relativamente estrechos. La estratificación en grupos de edad de muy diferente amplitud (p.e. 0-45 y 46+ años) es poco eficaz ya que, dentro de cada estrato, todavía puede haber diferencias considerables en la distribución de edad de los casos y controles (en un estudio caso-control) o entre los expuestos y no expuestos (en un estudio de cohorte o de intervención).

14.4 Confusión e interacción

Una suposición subyacente en el cálculo de un estimador resumen de efecto es que el efecto real es el mismo en cada uno de los estratos y que toda desviación de este efecto uniforme es debida a la variación aleatoria de muestreo. Una variación importante entre las estimaciones del efecto específicas de los diferentes estratos, indica la presencia de una *interacción* (también denominada *modificación del efecto*) entre la exposición de interés y el supuesto factor de confusión.

Ejemplo 14.4. Supongamos que estamos interesados en examinar la relación entre una exposición A y una enfermedad B en un estudio de cohorte. Empezamos calculando la razón de tasas cruda, que proporciona un valor de 2,0. A continuación decidimos examinar la relación entre A y B por separado para los expuestos a una tercera variable C (estrato 1) y los no expuestos (estrato 2). La Tabla 14.10 muestra tres resultados posibles de este estudio y el modo de interpretarlos. En la situación I, no existe confusión, debido a que las razones de tasas cruda y ajustada son similares, ni interacción, debido a que las razones de tasas son similares para los dos estratos. En la situación II, existe confusión, debido a que las razones de tasas cruda y ajustada son diferentes, pero no interacción, porque el efecto es similar en los dos estratos. En la situación III, existe una fuerte interacción entre A y C porque las razones de tasas específicas de estrato son considerablemente diferentes para los expuestos y los no expuestos a C.

	Razón de tasas cruda	Razón de tasas en el estrato 1	Razón de tasas en el estrato 2	Razón de tasas ajustada	
Situación I	2,0	2,0	2,0	2,0	No confusión No interacción
Situación II	2,0	3,0	3,0	3,0	Confusión presente No interacción
Situación III	2,0	4,0	0,5	–	Fuerte interacción

Tabla 14.10.
Ejemplo de confusión e interacción.

Si existe interacción entre las exposiciones en estudio y el factor de confusión, el cálculo de una estimación resumen del efecto de Mantel-Haenszel es erróneo, ya que no es una buena descripción de la asociación exposición-enfermedad, que en este caso, varía según el nivel de la variable de estratificación. Por tanto, sólo debemos calcular una estimación resumen ajustada del efecto, tomando una media ponderada de los estimadores de efecto de cada estrato, si estamos seguros de que las medidas del efecto específicas de estrato no varían entre sí (es decir, no existe interacción). Este concepto no es nuevo. Se comentó brevemente en el Apartado 4.3.3 al mencionar que las tasas estandarizadas por edad eran un resumen adecuado de los datos sólo si el efecto era homogéneo en los estratos de edad.

El primer paso en todo análisis estratificado es evaluar si existe interacción (Ejemplo 14.4). En la mayoría de los casos, esta decisión debe basarse en la inspección visual de los datos al examinar los patrones de variación en los estimadores específicos de cada estrato. Si son iguales, no existe interacción. No obstante, en esta situación puede producirse confusión si los efectos específicos de estrato difieren del efecto crudo. Si los estimadores específicos de estrato difieren entre sí, es indicativo de que existe interacción y en el análisis se debe prestar especial atención en describir como se modifica la asociación de interés en relación al factor de estratificación. En este caso, todos los estimadores de efecto específicos de estrato, y sus intervalos de confianza, deben indicarse por separado.

Con frecuencia, en un análisis en particular es difícil decidir si existe interacción o no porque los estimadores específicos de estrato varían, incluso aunque el efecto real sea similar, debido a la variación aleatoria. Por ejemplo, el efecto del frotis de Papanicolaou sobre la incidencia del cáncer de cuello uterino en el Ejemplo 14.3 era más marcado en mujeres con nivel educativo bajo (RT = 0,20; IC 95% = 0,08-0,54) que en las mujeres con nivel educativo alto (RT = 0,40; IC 95% = 0,23-0,69), indicando una interacción débil entre el frotis de VPH y el nivel educativo. No obstante, esta diferencia entre las razones de tasas específicas de estrato puede reflejar sólo una variación aleatoria. Existen numerosos tests de χ^2 de heterogeneidad para probar la hipótesis nula de que el grado de variabilidad en la serie de estimadores específicos de estrato se corresponde con la variación aleatoria. No obstante, en la práctica, estos tests no son muy potentes. Los estudios epidemiológicos, a no ser que estén diseñados específicamente para ello, raramente tienen el suficiente poder estadístico para detectar interacciones (véase Apartado 15.4.4). Por tanto, normalmente se adelanta más por inspección visual del tamaño y patrón de las estimaciones del efecto en los estratos que mediante los tests de interacción (modificación del efecto).

14.5 Uso del método de Mantel-Haenszel para ajustar por varios factores de confusión

En la práctica, a menudo existe la necesidad de examinar y posiblemente ajustar por varios factores de confusión. Esto puede conseguirse utilizando el método de Mantel-Haenszel aunque, a medida que aumenta el número de factores de confusión, este método se vuelve poco práctico porque la mayoría de los estratos tendrán datos muy escasos. Como veremos más adelante en este capítulo, los modelos de regresión son métodos más eficaces en esas circunstancias.

En el Ejemplo 14.5, la *odds ratio* cruda es 1,54. El intervalo de confianza al 95% y el valor de *P* indican que el azar es una explicación poco probable de esta observación. Por tanto, las mujeres que fumaban presentaban una mayor probabilidad de ser VPH positivas que las que no fumaban.

El número de parejas sexuales es un factor de riesgo conocido para la infección por VPH y puede haber confundido la asociación entre fumar y la

Ejemplo 14.5. Supongamos que se realizó un estudio caso-control para examinar los factores de riesgo de infección del cuello uterino por el virus del papiloma humano (VPH). Se utilizó un cuestionario estándar para obtener información detallada sobre las variables sociodemográficas y el comportamiento sexual de 188 mujeres VPH positivas (casos) y 571 mujeres VPH negativas (controles). La Tabla 14.11 muestra la distribución de casos y controles por hábito de fumar e infección por VPH.

	Fumar		Total
	Sí	No	
VPH-positiva	42 (a)	146 (b)	188 (n ₁)
VPH-negativa	90 (c)	481 (d)	571 (n ₀)
Total	132 (m₁)	627 (m₀)	759 (N)

Odds ratio cruda = (42/146) / (90/481) = 1,54
 Intervalo de confianza al 95% = 1,02–2,32
 Test $\chi^2 = 4,25$ con 1 g.l.; P = 0,039

Tabla 14.11.

Resultados de un estudio caso-control sobre fumar e infección por VPH: datos hipotéticos.

< 2 parejas sexuales			
	Fumar		Total
	Sí	No	
VPH-positiva	31 (a ₁)	124 (b ₁)	155 (n ₁₁)
VPH-negativa	78 (c ₁)	437 (d ₁)	515 (n ₀₁)
Total	109 (m₁₁)	561 (m₀₁)	670 (N₁)

Odds ratio = 1,40
 Intervalo de confianza al 95% = 0,88–2,22

≥ 2 parejas sexuales			
	Fumar		Total
	Sí	No	
VPH-positiva	11 (a ₂)	22 (b ₂)	33 (n ₁₂)
VPH-negativa	12 (c ₂)	44 (d ₂)	56 (n ₀₂)
Total	23 (m₁₂)	66 (m₀₂)	89 (N₂)

Odds ratio = 1,83
 Intervalo de confianza al 95% = 0,70–4,80

Test χ^2 de heterogeneidad = 0,24 con 1 g.l.; P = 0,63

$$OR_{MH} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i} = \frac{(31 \times 437) / 670 + (11 \times 44) / 89}{(124 \times 78) / 670 + (22 \times 12) / 89} = \frac{25,66}{17,40} = 1.47$$

Intervalo de confianza al 95% = 0,95–2,26
 $\chi^2 = 2,97$ con 1 g.l.; P = 0,09

Tabla 14.12.

Estudio caso-control hipotético sobre fumar e infección por VPH descrito en el Ejemplo 14.5: resultados estratificados por número de parejas sexuales en la vida.

infección por VPH. Las Tabla 14.12 muestra la asociación encontrada entre fumar y la infección por VPH, estratificada por el número referido de parejas sexuales a lo largo de la vida (clasificadas según <2 parejas y ≥ 2 parejas).

El examen de las *odds ratios* específicas de estrato (y sus intervalos de confianza) indica que el efecto de fumar sobre la infección por VPH en mujeres que informaron menos de dos parejas sexuales es similar al efecto en las que indicaron dos o más. Esto se confirma por el test de χ^2 de heterogeneidad. Debido a que el efecto de fumar sobre la infección por VPH es uniforme en los dos estratos, es apropiado calcular una *odds ratio* ajustada global. Por tanto, tras ajustar por el número de parejas sexuales, la *odds ratio* de fumar se reduce un poco, de 1,54 a 1,47. Este resultado indica que esta variable era un factor de confusión débil de la asociación entre fumar e infección por VPH. Pero incluso tras ajustar por el número de parejas sexuales, el intervalo de confianza al 95% sigue indicando un efecto de fumar sobre la infección por VPH. En resumen, estos resultados indican que el fumar aumentaba el riesgo de infección por VPH, pero este riesgo era marginalmente más débil tras considerar el efecto de confusión del número de parejas sexuales.

La misma técnica puede utilizarse para obtener una *odds ratio* ajustada por la edad. En este estudio hipotético, la edad (en años) se clasificó en 6 grupos (<20 , 20-24, 25-29, 30-34, 35-44, ≥ 45). De este modo, era necesario construir seis tablas 2×2 de fumar e infección por VPH, una para cada grupo de edad. Las celdas de estas tablas 2×2 se muestran en la Tabla 14.13.

Tabla 14.13.

Estudio caso-control hipotético sobre fumar e infección por VPH descrito en el Ejemplo 14.5: resultados estratificados por grupo de edad.

Estrato	Edad (años)	a_i	b_i	c_i	d_i	N_i	$a_i d_i / N_i$	$b_i c_i / N_i$
1	<20	3	10	16	79	108	2,19	1,48
2	20-24	13	44	16	92	165	7,25	4,27
3	25-29	6	33	22	62	123	3,02	5,90
4	30-34	8	25	10	75	118	5,08	2,12
5	35-44	8	22	18	89	137	5,20	2,89
6	≥ 45	4	12	8	84	108	3,11	0,89
Todos los estratos		42	146	90	481	759	25,85	17,55

Test χ^2 de heterogeneidad = 7,37 con 5 g.l.; $P = 0,20$

$$OR_{MH} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i} = \frac{25,85}{17,55} = 1,47$$

Intervalo de confianza al 95% = 0,96 – 2,32
 Test $\chi^2 = 3,06$ con 1 g.l.; $P = 0,08$

La OR_{MH} ajustada por la edad es ligeramente menor que la *odds ratio* cruda (1,47 frente a 1,54), indicando que la edad también era una variable de confusión débil de la asociación entre fumar e infección por VPH.

Así, hasta ahora, hemos estimado el efecto de fumar sobre la infección por VPH ajustado por el número de parejas sexuales y ajustado por la edad.

Podríamos estimar el efecto de fumar sobre la infección por VPH ajustado *simultáneamente* según el número de parejas y la edad utilizando el método de Mantel-Haenszel. Para hacer esto, es necesario construir una tabla 2 × 2 de fumar por infección por VPH para cada combinación posible del número de parejas y grupo de edad. Debido a que el número de parejas forma 2 estratos (<2 y ≥2) y la edad forma 6 estratos (<20, 20-24, 25-29, 30-34, 35-44, ≥45), es necesario construir 2 × 6 = 12 tablas 2 × 2. Las celdas de las tablas 2 × 2 para fumar para estos 12 estratos se muestran en la Tabla 14.14.

Estrato	Edad (años)	Número de parejas	a_i	b_i	c_i	d_i	N_i	$a_i d_i / N_i$	$b_i c_i / N_i$
1	<20	<2	2	8	11	67	88	1,52	1,00
2	20-24	<2	10	36	14	85	145	5,86	3,48
3	25-29	<2	3	28	19	59	109	1,62	4,88
4	30-34	<2	5	21	10	65	101	3,22	2,08
5	35-44	<2	7	19	17	78	121	4,51	2,67
6	≥45	<2	4	12	7	83	106	3,13	0,79
7	<20	≥2	1	2	5	12	20	0,60	0,50
8	20-24	≥2	3	8	2	7	20	1,05	0,80
9	25-29	≥2	3	5	3	3	14	0,64	1,07
10	30-34	≥2	3	4	0	10	17	1,76	0,00
11	35-44	≥2	1	3	1	11	16	0,69	0,19
12	≥45	≥2	0	0	1	1	2	0,00	0,00
All strata			42	146	90	481	759	24,60	17,46

Test χ^2 de heterogeneidad = 12,44 con 10 g.l.; $P = 0,26$

$$OR_{MH} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i} = \frac{24,60}{17,46} = 1,41$$

Intervalo de confianza al 95% = 0,91-2,24
 Test $\chi^2 = 2,62$ con 1 g.l.; $P = 0,132$

Tabla 14.14. Estudio caso-control hipotético sobre fumar e infección por VPH descrito en el Ejemplo 14.5: resultados estratificados por la edad y el número de parejas sexuales en la vida.

Por tanto, tras ajustar según los efectos de confusión del número de parejas y la edad, el efecto de fumar sobre la infección por VPH es todavía más pequeño ($OR_{MH} = 1,41$).

El método de Mantel-Haenszel puede ampliarse para ajustar simultáneamente por más de dos variables de confusión. Por ejemplo, para estimar el efecto de fumar sobre la infección por VPH, según el número de parejas sexuales (dos estratos), la edad (seis estratos), el estado civil (tres estratos: casada, soltera, viuda/divorciada) y el nivel educativo (tres estratos), construiríamos $2 \times 6 \times 3 \times 3 = 108$ tablas 2 × 2. Obviamente, 108 tablas 2 × 2 formadas de una serie de datos de 188 casos y 571 controles significa que algunos estratos tendrán un número de observaciones muy pequeño, o ninguno.

Otro problema con el método de Mantel-Haenszel es que cada variable explicativa incluida en el análisis tiene que clasificarse como una exposi-

ción o como una variable de confusión, y puede haber sólo una exposición. Por ejemplo, fumar era nuestra exposición, y el número de parejas y la edad eran nuestras variables de confusión. Por tanto obtuvimos una *odds ratio* para fumar ajustada por el número de parejas y edad. Estos resultados no nos proporcionan la *odds ratio* para el número de parejas ajustada por el fumar y la edad, o la *odds ratio* para edad ajustada por fumar y el número de parejas. Éstas requerirían análisis adicionales de Mantel-Haenszel.

14.6 Uso de modelos de regresión para ajustar por el efecto de los factores de confusión

Los modelos de regresión resumen la relación entre una variable resultado (también denominada dependiente) y varias variables explicativas (independientes) en forma de una ecuación matemática. La forma general de esta ecuación es

Variable resultado = función (variables explicativas)

Existen muchos tipos de modelos de regresión. La elección de un modelo en particular depende de las características de la variable resultado (por ejemplo, continua o categórica) y de la forma en que está relacionada matemáticamente con las variables explicativas. El modelo matemático más simple que podríamos utilizar ya se ha introducido en el Apartado 11.2.1 para describir la relación entre dos variables cuantitativas.

La discusión de estos modelos y las suposiciones en las que se basan está fuera del alcance de este capítulo. Estas técnicas estadísticas se utilizan habitualmente en estudios epidemiológicos y, por tanto, en el resto de este capítulo intentaremos ilustrar cómo están relacionadas con el método de Mantel-Haenszel para que el lector pueda entender e interpretar los resultados de trabajos publicados en los que hayan sido utilizados estos métodos.

Volvamos a considerar el estudio caso-control hipotético descrito en el Ejemplo 14.5. Podemos utilizar una técnica de regresión en particular, denominada *regresión logística*, para analizar los datos de estudios caso-control no apareados. En este análisis, empezamos utilizando un modelo de regresión logística que incluye como variables explicativas sólo la exposición en estudio –fumar. Los resultados se muestran en la Tabla 14.15 (modelo 1). La *odds ratio* estimada por este modelo de regresión logística es la misma que la *odds ratio* cruda obtenida en la Tabla 14.11.

A continuación utilizamos un modelo que incluye el fumar y el número de parejas sexuales como variables explicativas (modelo 2 en la Tabla 14.15). Este modelo proporciona

Odds ratio para fumar frente a no fumar ajustada por número de parejas sexuales = 1,47 (IC 95% = 0,97 a 2,23)

De este modo, tras ajustar según el número de parejas sexuales, el efecto de fumar sobre la infección por VPH disminuyó (1,47 frente a 1,54). Este resultado es similar al obtenido anteriormente cuando utilizamos la técnica de Mantel-Haenszel para controlar por el efecto del número de parejas sexuales (Tabla 14.12). Pero a diferencia de esta técnica, este modelo de regresión también nos proporciona la *odds ratio* para el número de parejas sexuales ajustada por el fumar:

Odds ratio para ≥ 2 parejas sexuales frente a < 2 , ajustada por fumar
= 1,90 (IC 95% = 1,19 a 3,03)

Así, existe un riesgo aumentado, estadísticamente significativo, de infección por VPH asociado con dos o más parejas sexuales incluso después de tener en cuenta las diferencias en fumar.

Podemos utilizar un modelo más complejo que incluye el fumar, el número de parejas sexuales y la edad (modelo 3, Tabla 14.15).

Variabes	Odds ratio	Intervalo de confianza al 95%
<i>Modelo 1</i>		
fumar ^a	<u>1,54</u>	1,02–2,32
<i>Modelo 2</i>		
fumar ^a	<u>1,47</u>	0,97–2,23
parejas ^b	1,90	1,19–3,03
<i>Modelo 3</i>		
fumar ^a	<u>1,43</u>	0,93–2,19
parejas ^b	1,95	1,19–3,17
edad 2 ^c	4,14	2,12–8,12
edad 3	3,58	1,77–7,24
edad 4	3,01	1,47–6,15
edad 5	2,18	1,07–4,46
edad 6	1,49	0,67–3,33

^a categorizada como 'no fumadores' (basal) y 'fumadores'.

^b categorizada como '< 2 parejas' (basal) y '≥ 2 parejas'.

^c categorizada como edad 1 = <20 (basal); edad 2 = 20–24; edad 3 = 25–29; edad 4 = 30–34; edad 5 = 35–44; edad 6 = 45+ años.

Tabla 14.15.

Estudio caso-control hipotético sobre los factores de riesgo de infección de cuello uterino por VPH descrito en el Ejemplo 14.5. Resultados obtenidos a partir de modelos de regresión logística que incluían un número creciente de variables explicativas: fumar, número de parejas sexuales en la vida y edad. (Los valores subrayados corresponden a los obtenidos anteriormente con la técnica de Mantel-Haenszel mostrada en las Tablas 14.11, 14.12 y 14.14).

Este último modelo nos proporciona una estimación de 1,43 (IC 95% = 0,93 – 2,19), que es similar a la obtenida anteriormente con el método de Mantel-Haenszel ($OR_{MH} = 1,41$; IC 95% = 0,91 – 2,24) (véase Tabla 14.14). No obstante, a diferencia del análisis de Mantel-Haenszel, el análisis de regresión logística también nos proporciona las siguientes *odds ratios*

Odds ratio para ≥ 2 parejas sexuales frente a < 2 , ajustada por fumar y edad = 1,95 (IC 95% = 1,19 – 3,17)

Odds ratio para edad 20-24 frente a la edad <20, ajustada por fumar y parejas = 4,14 (IC 95% = 2,12 – 8,12)

Odds ratio para edad 25-29 frente a la edad <20, ajustada por fumar y parejas = 3,58 (IC 95% = 1,77 – 7,24)

Odds ratio para edad 30-34 frente a la edad <20, ajustada por fumar y parejas = 3,01 (IC 95% = 1,47 – 6,15)

Odds ratio para edad 35-44 frente a la edad <20, ajustada por fumar y parejas = 2,18 (IC 95% = 1,07 – 4,46)

Odds ratio para edad 45+ frente a la edad <20, ajustada por fumar y parejas = 1,49 (IC 95% = 0,67 – 3,33)

Una de las principales ventajas de los modelos de regresión es que no requieren la definición de qué variable explicativa es la exposición y cuáles son las variables de confusión potenciales, ya que todas las variables explicativas se tratan del mismo modo. Esto es especialmente importante en estudios diseñados para examinar el efecto de una gran cantidad de exposiciones en lugar del efecto de una exposición específica única.

Modelos de regresión similares pueden aplicarse a los datos de otros diseños de estudios. Volvamos a considerar el estudio de cohorte hipotético sobre el frotis de Papanicolaou y el cáncer de cuello uterino descrito en el Ejemplo 14.3 (Apartado 14.2.3). En la Tabla 14.8 calculamos la razón de tasas cruda para medir el efecto del frotis de Papanicolaou sobre el cáncer de cuello uterino. A continuación seguimos calculando el efecto del frotis de Papanicolaou ajustando por el nivel educativo (Tabla 14.9) utilizando la técnica de Mantel-Haenszel. También podemos utilizar la técnica de Mantel-Haenszel para ajustar simultáneamente según el nivel educativo, estado civil y edad en la primera relación. Los resultados se muestran en la Tabla 14.16.

Tabla 14.16. Estudio de cohorte hipotético sobre el frotis de Papanicolaou y el cáncer de cuello uterino descrito en el Ejemplo 14.3. Resultados obtenidos mediante la técnica de Mantel-Haenszel.

Variables por las que se ha ajustado	Razón de tasas de cáncer de cuello uterino para frotis de Papanicolaou	Intervalo de confianza al 95%
Ninguna	0,32 (cruda)	0,20–0,52
Nivel educativo ^a	0,32	0,20–0,52
Nivel educativo y estado civil ^b	0,40	0,25–0,66
Nivel educativo, estado civil y edad al primer coito ^c	0,43	0,27–0,72

^a Clasificado según 'nivel educativo bajo' y 'nivel educativo alto'.

^b Clasificado según estado civil 1=casada; 2=soltera; 3=divorciada/viuda.

^c Clasificado según edad al primer coito 1=<18 años; 2=18–22 años; 3=22+ años.

La utilización de la técnica de Mantel-Haenszel para ajustar simultáneamente por el nivel educativo, estado civil y edad en la primera relación im-

plicó la formación de 18 ($= 2 \times 3 \times 3$) estratos. No obstante, sólo se produjeron diecisiete casos de cáncer de cuello uterino en mujeres que indicaron haberse realizado alguna vez un frotis de Papanicolaou. Por consiguiente, había celdas vacías en varios estratos.

Podemos utilizar una técnica especial de regresión, denominada *regresión de Poisson*, para analizar los datos de este estudio de cohorte hipotético sobre el frotis de Papanicolaou y el cáncer de cuello uterino. Los resultados se muestran en la Tabla 14.17.

Variable	Tasa basal (por persona y año)	Razón de tasas de cáncer de cuello uterino	Intervalo de confianza al 95%
Modelo 1			
constante	0,00074		0,0007–0,0008
Frotis de Papanicolaou ^a		<u>0,32</u>	0,20–0,52
Modelo 2			
constante	0,008		0,008–0,009
Frotis de Papanicolaou ^a		<u>0,32</u>	0,20–0,52
Nivel educativo ^b		0,73	0,65–0,82
Modelo 3			
constante	0,005		0,004–0,005
Frotis de Papanicolaou ^a		<u>0,41</u>	0,25–0,66
Nivel educativo ^b		0,74	0,66–0,84
Estado civil ^{2c}		2,68	2,28–3,15
Estado civil3		1,89	1,61–2,21
Modelo 4			
constante	0,008		0,006–0,009
Frotis de Papanicolaou ^a		<u>0,46</u>	0,29–0,75
Nivel educativo ^b		0,77	0,68–0,87
Estado civil ^{2c}		2,68	2,27–3,15
Estado civil3		1,60	1,36–1,87
Edad al primer coito ^{2d}		0,52	0,46–0,59
Edad al primer coito3		0,13	0,09–0,19

^a Clasificado como 'nunca' (basal) y 'alguna vez'.

^b Clasificado como 'nivel educativo bajo' (basal) y 'nivel educativo alto'.

^c Clasificado como estado civil 1=casada (basal), 2=soltera, 3=divorciada/viuda.

^d Clasificado según edad al primer coito 1= <18 años(basal), 2=18–22 años, 3=22+ años.

Tabla 14.17.

Estudio de cohorte hipotético sobre el frotis de Papanicolaou y el cáncer de cuello uterino descrito en el Ejemplo 14.3. Resultados obtenidos a partir de modelos de regresión de Poisson con números crecientes de variables explicativas. (Los valores subrayados corresponden a los obtenidos con la técnica de Mantel-Haenszel mostrada en la Tabla 14.16).

Al igual que con la regresión logística, empezamos utilizando un modelo que incluye sólo una variable explicativa –el frotis de Papanicolaou (modelo 1 en la Tabla 14.17). Este modelo nos proporciona la estimación de Poisson de la razón de tasas de cáncer de cuello uterino para el frotis de Papanicolaou. Este valor corresponde a la razón de tasas cruda obtenida ante-

riormente. Este modelo nos proporciona también otro valor denominado 'constante', que corresponde a la tasa de incidencia de cáncer de cuello uterino en mujeres que indicaron no haberse realizado nunca un frotis de Papanicolaou (véase la Tabla 14.8). A partir de estos valores, podemos calcular la tasa de incidencia en los expuestos (mujeres que indicaron haberse realizado por lo menos un frotis de Papanicolaou) como $0,00074 \times 0,32 = 0,0002368 = 24$ por 100.000 personas-años (el mismo valor obtenido en la Tabla 14.8).

Después se continúa añadiendo a nuestro modelo otra variable explicativa, por ejemplo, el nivel educativo (modelo 2 en la Tabla 14.17). Este modelo nos proporciona la estimación de Poisson de la razón de tasas de cáncer de cuello uterino para frotis de Papanicolaou ajustada por el nivel educativo, que es 0,32 (IC 95% = 0,20 – 0,52). Este es el mismo valor obtenido al utilizar la técnica de Mantel-Haenszel para controlar por el nivel educativo (Tabla 14.16). A diferencia de la técnica de Mantel-Haenszel, este modelo también nos proporciona

Razón de tasas de cáncer de cuello uterino entre un nivel educativo alto y un nivel educativo bajo, ajustada por el uso de frotis de Papanicolaou = 0,73 (IC 95% = 0,65 – 0,82)

En el último modelo (modelo 4) mostrado en la Tabla 14.17, incluimos frotis de Papanicolaou, nivel educativo, estado civil y edad a la primera relación sexual como variables explicativas. La estimación de Poisson de la razón de tasas de cáncer de cuello uterino para frotis de Papanicolaou ajustada por el nivel educativo, estado civil y edad a la primera relación es 0,46 (IC 95% = 0,29 – 0,75), similar a la obtenida con el método de Mantel-Haenszel ($RT_{MH} = 0,43$; IC 95% = 0,27 – 0,72) (Tabla 14.16). Pero con la regresión de Poisson obtuvimos también las siguientes razones de tasas adicionales:

Razón de tasas de cáncer de cuello uterino para un nivel educativo alto frente a uno bajo ajustada por frotis de Papanicolaou, estado civil y edad a la primera relación sexual = 0,77 (IC 95% = 0,68 – 0,87)

Razón de tasas de cáncer de cuello uterino para mujeres solteras frente a casadas ajustada para frotis de Papanicolaou, nivel educativo y edad a la primera relación sexual = 2,68 (IC 95% = 2,27 – 3,15)

Razón de tasas de cáncer de cuello uterino para mujeres divorciadas/viudas frente a casadas ajustada por frotis de Papanicolaou, nivel educativo y edad a la primera relación sexual = 1,60 (IC 95% = 1,36 – 1,87)

Razón de tasas de cáncer de cuello uterino para edad a la primera relación entre 18-22 frente a <18 años ajustada por frotis de Papanicolaou, nivel educativo y estado civil = 0,52 (IC 95% = 0,46 – 0,59).

Razón de tasas de cáncer de cuello uterino para edad a la primera relación sexual a 22+ frente a <18 años ajustada por frotis de Papanicolaou, nivel educativo y estado civil = 0,13 (IC 95% = 0,09 – 0,19).

La regresión logística se utiliza para estimar *odds ratios*. Por tanto puede utilizarse en un estudio caso-control, un estudio transversal o, si se estiman «riesgos» en lugar de «tasas», en un estudio de cohorte. Los modelos de regresión de Poisson se utilizan para estimar razones de tasas utilizando datos de persona-tiempo. Otros modelos de regresión utilizados habitualmente en epidemiología son:

Regresión logística condicional – el análisis de regresión logística es adecuado para estudios caso-control no apareados o para estudios caso-control apareados por frecuencias. Los estudios caso-control apareados individualmente requieren una estrategia ligeramente diferente denominada análisis de regresión logística condicional. Esta técnica es la única forma en que puede ajustarse por los factores de confusión que no sean el(los) factor(es) de apareamiento utilizados en el diseño de estos estudios.

Modelo de riesgos proporcionales de Cox – este tipo de modelo de regresión se utiliza cuando el *tiempo transcurrido hasta un acontecimiento* es de especial interés (como es el caso en el análisis de la supervivencia).

14.7 Conclusiones

En resumen, el método de Mantel-Haenszel es una técnica muy útil para ajustar por factores de confusión, y esta estrategia es con frecuencia adecuada para datos con pocos factores de confusión. No obstante, con objeto de ajustar simultáneamente por varios factores de confusión puede ser necesario emplear modelos de regresión.

Es importante destacar que todo análisis debe empezar utilizando el método de Mantel-Haenszel en el análisis preliminar para obtener estimaciones del efecto crudas y estimaciones del efecto ajustadas para cada uno de los factores de confusión por separado. La técnica de Mantel-Haenszel emplea datos tabulados; por tanto, el investigador se mantiene en contacto con los datos ya que las tabulaciones cruzadas utilizadas para la estratificación le permiten observar la mayor parte de las relaciones e interacciones importantes presentes. Éstas permiten al investigador la posibilidad de detectar errores e incompatibilidades en los datos que de otro modo no se detectarían.

Los modelos de regresión pueden utilizarse a continuación, en una segunda fase del análisis, para ajustar simultáneamente por varios factores de confusión. Uno de los principales inconvenientes de los modelos de regresión es que se pierde la visión global de los datos, razón por la que con frecuencia se considera una estrategia de «caja negra». Las personas que no están familiarizadas con los modelos estadísticos y las que no entienden las asunciones en las que se basan, no los deben utilizar. El usuario debe entender estas asunciones porque pueden o no ser aplicables a las variables en estudio.

Lecturas adicionales

* En este capítulo se han presentado fórmulas para calcular las razones de tasas, razones de riesgos y *odds ratios* de Mantel-Haenszel. Las fórmulas para calcular las estimaciones ajustadas de diferencias de riesgos y de tasas se pueden encontrar en Greenland y Robins (1985).

* Las técnicas de estratificación y modelos de regresión se tratan de manera mucho más detallada (aunque más elaborada estadísticamente) en Breslow y Day (1980, 1987) y Clayton y Hills (1993).

Cuadro 14.1. Puntos esenciales

- Cualquier análisis de datos debe planearse detenidamente. En general, debe incluir los siguientes pasos:
 1. Producir tablas simples para comprobar la coherencia de los datos.
 2. Calcular medidas del efecto *crudas*.
 3. Estratificar por niveles del factor de confusión potencial.
 4. Calcular estimaciones del efecto específicas para cada estrato.
 5. Comprobar la uniformidad de las estimaciones específicas para cada estrato por inspección visual o realizando un test de significación estadística.
 6. Si se cree que el efecto es uniforme en los estratos, calcular una estimación resumen ajustada del efecto combinado utilizando el método de Mantel-Haenszel. Calcular los intervalos de confianza para la estimación ajustada y el test χ^2 de Mantel-Haenszel.
 7. Si se cree que el efecto no es uniforme (es decir, si existe interacción), indicar las estimaciones específicas de estrato, los intervalos de confianza y el test χ^2 para cada estimación.
 8. Utilizar técnicas de modelos de regresión para ajustar simultáneamente según varios factores de confusión.
- En la fase inicial de un análisis deben utilizarse *siempre* métodos clásicos simples basados en la estratificación. Las tabulaciones cruzadas utilizadas en la estratificación mantienen al investigador en contacto directo con los datos.
- Los modelos de regresión pueden utilizarse en una segunda fase del análisis para ajustar simultáneamente por varios factores de confusión. A diferencia de los métodos clásicos, los modelos de regresión son, en cierto grado, una estrategia de «caja negra» y por esta razón, pueden dar lugar a errores importantes. Estos métodos son procedimientos estadísticos complejos que nunca deben emplear los investigadores que no estén familiarizados con ellos.

Anexo 14.1.

Intervalos de confianza y tests estadísticos para las medidas de efecto ajustadas

En este capítulo hemos presentado fórmulas para calcular la razón de tasas, la razón de riesgos y la *odds ratio* de Mantel-Haenszel. Las fórmulas para calcular las estimaciones de las diferencias de tasas y diferencias de riesgo ajustadas pueden encontrarse en Greenland y Robins (1985). Observe que existen varios métodos para calcular un intervalo de confianza para la *odds ratio* ajustada, que pueden proporcionar valores ligeramente diferentes de los obtenidos aquí. Estos cálculos pueden realizarse fácilmente mediante paquetes estadísticos tales como EPI INFO, STATA o EGRET.

A14.1.1 *Odds ratio* ajustada

Intervalo de confianza para la *odds ratio* de Mantel-Haenszel

El error estándar (EE) del logaritmo de la *odds ratio* de Mantel-Haenszel (OR_{MH}) puede estimarse como

$$EE (\ln OR_{MH}) = \sqrt{\frac{\sum (b_i c_i / N_i)^2 v_i}{(\sum b_i c_i / N_i)^2}}$$

donde $\sum b_i c_i / N_i$ corresponde al denominador de la fórmula utilizada para calcular la OR_{MH} y $v_i = 1/a_i + 1/b_i + 1/c_i + 1/d_i$

Así, en el estudio de cáncer de ovario (Ejemplo 14.1; Tabla 14.2) tenemos

$$\sum b_i c_i / N_i = (32 \times 8) / 77 + (4 \times 50) / 81 = 3,32 + 2,47 = 5,79$$

$$v_1 = 1/9 + 1/32 + 1/8 + 1/28 = 0,30$$

$$v_2 = 1/15 + 1/4 + 1/50 + 1/12 = 0,42$$

$$\sum (b_i c_i / N_i)^2 v_i = (3,32^2 \times 0,30) + (2,47^2 \times 0,42) = 3,31 + 2,56 = 5,87$$

Por tanto,

$$EE (\ln OR_{MH}) = \sqrt{5,87 / (5,79)^2} = 0,42$$

A continuación puede calcularse un intervalo de confianza al 95% 'aproximado' para el $\ln OR_{MH}$ como

$$IC \ 95\% (\ln OR_{MH}) = (\ln 0,95) \pm 1,96 \times 0,42 = -0,05 \pm 1,96 \times 0,42 = -0,87 \text{ a } 0,77$$

Tomando antilogaritmos puede obtenerse un intervalo de confianza al 95% 'aproximado' para la OR_{MH}

$$IC\ 95\% (OR_{MH}) = e^{-0,87} \text{ a } e^{0,77} = 0,42 \text{ a } 2,16$$

Test de χ^2 de Mantel-Haenszel

El test de χ^2 de Mantel-Haenszel puede utilizarse para determinar si la OR_{MH} es significativamente diferente de uno. El test es sólo una extensión del test de χ^2 de Mantel-Haenszel para una única tabla 2×2 presentado en el Apartado A6.1.3. La hipótesis nula es que no existe una asociación entre la exposición y la enfermedad (es decir, la *odds ratio* es igual a uno) en cada uno de los estratos individuales. Con objeto de realizar este test, antes debe obtenerse lo siguiente para cada estrato i :

(i) El valor observado de a_i :

$$O(a_i) = a_i$$

(ii) El valor esperado de a_i , asumiendo la hipótesis nula de ausencia de asociación:

$$E(a_i) = n_{1i}m_{1i}/N_i$$

(iii) La varianza de a_i , asumiendo la hipótesis nula de ausencia de asociación:

$$V(a_i) = n_{1i}n_{0i}m_{1i}m_{0i} / (N_i^2(N_i - 1))$$

A continuación se suma cada una de estas cantidades en todos los estratos. En el Ejemplo 14.1. se obtiene:

$$\Sigma O(a_i) = 15 + 9 = 24$$

$$\Sigma E(a_i) = \Sigma n_{1i}m_{1i} / N_i = (41 \times 17)/77 + (19 \times 65)/81 = 9,05 + 15,25 = 24,30$$

$$\Sigma V(a_i) = \Sigma n_{1i}n_{0i}m_{1i}m_{0i} / N_i^2 (N_i - 1) = (41 \times 36 \times 17 \times 60) / (77^2 \times 76) + (19 \times 62 \times 65 \times 16) / (81^2 \times 80) = 3,34 + 2,33 = 5,67$$

Sería de esperar que la diferencia entre los valores observados y esperados fuera pequeña si la hipótesis nula fuera cierta. Para probar si las diferencias obtenidas son mayores de las esperadas por azar, calculamos

$$\chi^2 = (\Sigma O(a_i) - \Sigma E(a_i))^2 / \Sigma V(a_i)$$

y obtenemos un valor P refiriendo nuestro resultado a la distribución de χ^2 con 1 grado de libertad (g.l.).

En nuestro ejemplo, $\chi^2 = (24 - 24,30)^2 / 5,67 = 0,016$ con 1 g.l. Esto da $P = 0,93$ a partir de la cual se concluye que tras ajustar según el uso de anti-conceptivos orales, no existen evidencias de asociación entre fumar y cáncer de ovario. Así, $OR_{MH} = 0,95$ no es significativamente diferente de 1.

A14.1.2 Razón de riesgos ajustada

Intervalo de confianza para la razón de riesgos de Mantel-Haenszel

El error estándar (EE) del logaritmo de la razón de riesgos de Mantel-Haenszel (R_{MH}) puede estimarse por

$$EE(\ln RR_{MH}) = \sqrt{\frac{\sum(n_{1i}m_{1i}m_{0i} - a_i b_i N_i) / N_i^2}{(\sum a_i m_{0i} / N_i)(\sum b_i m_{1i} / N_i)}}$$

donde las expresiones $\sum a_i m_{0i} / N_i$ y $\sum b_i m_{1i} / N_i$ corresponden, respectivamente, al numerador y denominador de la fórmula utilizada para calcular R_{MH} .

En nuestro estudio ocupacional (Ejemplo 14.2; Tabla 14.6)

$$\sum a_i m_{0i} / N_i = 264$$

$$\sum b_i m_{1i} / N_i = 132$$

$$\begin{aligned} \sum(n_{1i}m_{1i}m_{0i} - a_i b_i N_i) / N_i^2 &= (240 \times 4.000 \times 16.000 - 80 \times 160 \times 20.000) / \\ &20.000^2 + (600 \times 80.000 \times 80.000 - 400 \times 200 \\ &\times 160.000) / 160.000^2 = 37,76 + 149,50 = 187,26 \end{aligned}$$

Por tanto,

$$EE(\ln RR_{MH}) = \sqrt{187,26 / (264 \times 132)} = 0,073$$

Intervalo de confianza al 95% ($\ln RR_{MH}$) = $(\ln 2,0) \pm 1,96 \times 0,073 = 0,547$ a $0,833$

Tomando antilogaritmos puede obtenerse un intervalo de confianza al 95% 'aproximado' para la RR_{MH}

$$\text{Intervalo de confianza al 95\% } (RR_{MH}) = e^{0,547} \text{ a } e^{0,833} = 1,73 \text{ a } 2,30$$

Test de χ^2 de Mantel-Haenszel

El test de χ^2 de Mantel-Haenszel para evaluar si la RR_{MH} es significativamente diferente de la unidad es similar al utilizado anteriormente para la *odds ratio*. La hipótesis nula es que no existe una asociación entre la exposición y la enfermedad en cada uno de los estratos individuales, es decir, que la razón de riesgos es igual a uno en cada estrato.

En nuestro estudio ocupacional (Ejemplo 14.2; Tabla 14.6)

$$\sum O(a_i) = 80 + 400 = 480$$

$$\Sigma E(a_i) = \Sigma n_{1i} m_{1i} / N_i = 240 \times 4.000 / 20.000 + 600 \times 80.000 / 160.000 = 48 + 300 = 348$$

$$\Sigma V(a_i) = \Sigma n_{1i} n_{0i} m_{1i} m_{0i} / (N_i^2 (N_i - 1)) = (240 \times 19.760 \times 4.000 \times 16.000) / (20.000^2 \times 19.999) + (600 \times 159.400 \times 80.000 \times 80.000) / (160.000^2 \times 159.999) = 37,94 + 149,44 = 187,38$$

Ahora podemos calcular el test χ^2

$$\chi^2 = (\Sigma O(a_i) - \Sigma E(a_i))^2 / \Sigma V(a_i) = (480 - 348)^2 / 187,38 = 92,99$$

y obtener un valor P refiriendo nuestro resultado a la distribución χ^2 con 1 grado de libertad. En este ejemplo, $P < 0,0001$. Por tanto, es muy poco probable que estos resultados sean debidos al azar.

A14.1.3 Razón de tasas ajustada

Intervalo de confianza para la razón de tasas de Mantel-Haenszel

Al igual que con cualquier otro estimador, es útil poder construir un intervalo de confianza al 95% alrededor de una razón de tasas de Mantel-Haenszel (RT_{MH}). El error estándar del logaritmo de una razón de tasas puede estimarse por

$$EE(\ln RT_{MH}) = \sqrt{\frac{\Sigma V(a_i)}{(\Sigma a_i \gamma_{0i} / \gamma_i)(\Sigma b_i \gamma_{1i} / \gamma_i)}}$$

Podemos obtener un intervalo de confianza al 95% 'aproximado' alrededor del logaritmo de la RT_{MH} como

$$\ln RT_{MH} \pm 1,96 \times EE(\ln RT_{MH})$$

A continuación puede obtenerse un intervalo de confianza al 95% 'aproximado' para la RT_{MH} tomando antilogaritmos.

Observe que $\Sigma a_i \gamma_{0i} / \gamma_i$ y $\Sigma b_i \gamma_{1i} / \gamma_i$ son, respectivamente, el numerador y el denominador de la fórmula para la razón de tasas de Mantel-Haenszel, que se calcularon para el Ejemplo 14.3 en el Apartado 14.2.3 Por tanto, sólo necesitamos calcular $\Sigma V(a_i)$.

$$\Sigma a_i \gamma_{0i} / \gamma_i = 16,24$$

$$\Sigma b_i \gamma_{1i} / \gamma_i = 50,23$$

$$V(a_i) = \Sigma n_i \gamma_{1i} \gamma_{0i} / \gamma_i^2 = (710 \times 38.346 \times 828.149) / 866.495^2 + (431 \times 32.838 \times 690.552) / 723.390^2 = 30,03 + 18,68 = 48,71$$

$$\ln RT_{MH} = \ln(0,32) = -1,14$$

$$EE(\ln RT_{MH}) = \sqrt{48,71 / (16,24 \times 50,23)} = 0,244$$

$$\begin{aligned} \text{Intervalo de confianza al 95\% } (\ln RT_{MH}) &= -1,14 \pm 1,96 \times 0,244 \\ &= -1,62 \text{ a } -0,66 \end{aligned}$$

$$\text{Intervalo de confianza al 95\% } (RT_{MH}) = e^{-1,62} \text{ a } e^{-0,66} = 0,20 \text{ a } 0,52$$

Test de χ^2 de Mantel-Haenszel

Podemos realizar un test de significación para demostrar que la razón de tasas real es diferente de uno. Nuestra hipótesis nula es que la razón de tasas real en *todos* los estratos es uno. Para calcular el test, en primer lugar es necesario calcular lo siguiente de cada estrato

- (i) Valor observado de $a_i = O(a_i) = a_i$
- (ii) Valor esperado de $a_i = E(a_i) = n_i \gamma_{1i} / \gamma_i$
- (iii) Varianza de $a_i = V(a_i) = n_i \gamma_{1i} \gamma_{0i} / \gamma_i^2$

Un test de significación global (que la razón de tasas común es la unidad) viene dado por

$$\chi^2 = (\Sigma O(a_i) - \Sigma E(a_i))^2 / \Sigma V(a_i)$$

donde la función sumatoria es sobre todos los estratos. El valor calculado debe mirarse en las tablas de la distribución de χ^2 con un grado de libertad.

Por tanto, en el Ejemplo 14.3,

$$\Sigma O(a_i) = 13 + 4 = 17$$

$$\begin{aligned} \Sigma E(a_i) &= (710 \times 38.346 / 866.495) + (431 \times 32.838 / 723.390) = 31,42 + \\ &19,57 = 50,99 \end{aligned}$$

$\Sigma V(a_i) = 48,71$, obtenida anteriormente para el cálculo del intervalo de confianza.

$$\text{Test } \chi^2 = (17 - 50,99)^2 / 48,71 = 23,72$$

Esta χ^2 con 1 g.l. corresponde a una $P < 0,001$.

Capítulo 15

Tamaño de un estudio

15.1 Introducción

Es importante asegurar en la fase de diseño de cualquier tipo de estudio que el número de sujetos propuesto para incluir sea adecuado para responder al(los) objetivo(s) principal(es) del estudio. Un estudio pequeño puede dejar de detectar efectos importantes sobre los resultados de interés, o puede estimarlos de manera demasiado imprecisa, independientemente de la calidad del diseño. Un estudio más grande de lo necesario, aunque en la práctica es poco frecuente, puede desperdiciar recursos valiosos.

Ejemplo 15.1. Supongamos que se realizó un estudio para evaluar el valor de un tratamiento nuevo para el cáncer de mama. Un total de 200 mujeres con cáncer de mama recién diagnosticado se distribuyeron aleatoriamente para recibir el tratamiento nuevo o el estándar. Todas las pacientes se siguieron durante un año después de su inclusión en el estudio (o hasta la muerte si esta se producía antes). El resultado de interés era la proporción de mujeres todavía vivas al final del estudio. Los resultados se muestran en la Tabla 15.1.

		Tratamiento nuevo	Tratamiento estándar	Total
Vivas 1 año después de la inclusión en el estudio	Sí	80 (a)	70 (b)	150 (n_1)
	No	20 (c)	30 (d)	50 (n_0)
	Total	100 (m_1)	100 (m_0)	200 (N)

$$p_1 = 80/100 = 80\%$$

$$p_0 = 70/100 = 70\%$$

$$\text{Diferencia de riesgo} = p_1 - p_0 = 10\%$$

$$\text{Intervalo de confianza al 95\%} = -2\% \text{ a } +22\%$$

$$\text{Test } \chi^2 = 2,65 \text{ con 1 g.l.; } P > 0,10$$

Tabla 15.1.

Número de pacientes con cáncer de mama vivas un año después de la inclusión en el estudio según el tipo de tratamiento administrado: datos hipotéticos.

En el Ejemplo 15.1, el test de χ^2 sobre la significación de la diferencia entre estas dos proporciones proporciona un valor de 2,65, que corresponde a una $P > 0,10$. Por tanto la diferencia entre las dos proporciones podría haber aparecido fácilmente por azar. No obstante, a partir de estos datos no podemos concluir que no existe una diferencia real entre los tratamientos ya que el intervalo de confianza al 95% para la diferencia entre las proporcio-

nes de pacientes vivas 1 año después de la inclusión en el estudio es -2% a $+22\%$. Es decir, los datos de este estudio son compatibles a la vez con un aumento de la proporción de pacientes que sobreviven con el tratamiento nuevo de hasta el 22% ó con una reducción del 2% de la proporción de supervivientes entre las que recibieron el tratamiento estándar.

Por tanto, aunque el estudio ha mostrado que el tratamiento nuevo no funciona apreciablemente peor que el tratamiento estándar, no está claro si los dos tratamientos poseen efectos similares o si el tratamiento nuevo aumenta la supervivencia considerablemente. Esto es debido a que el tamaño de la muestra de este estudio era demasiado pequeño para proporcionar una respuesta adecuada a la cuestión tratada.

En este capítulo mostraremos como pueden obtenerse estimaciones del tamaño de la muestra necesario en la situación más simple en la que tienen que compararse dos grupos. Los cálculos se basan en los métodos estadísticos presentados en el Capítulo 6 y su anexo, y los lectores deben estar familiarizados con su contenido antes de continuar.

Existen dos métodos principales para los cálculos del tamaño de la muestra. Uno se basa en el concepto de *poder* estadístico de un estudio, es decir, su capacidad de detectar un resultado estadísticamente significativo si la magnitud real del efecto es como la que se ha previsto. Por tanto, esta aproximación a los cálculos del tamaño de la muestra se centra en el test de *significación* que se realizará al final del estudio. En el Ejemplo 15.1, podemos estimar el tamaño de la muestra necesario para asegurar que el estudio tendrá una cierta probabilidad ('poder') de proporcionar un valor de P inferior a $0,05$ si la diferencia real en la supervivencia entre los dos tratamientos es del 10% . El segundo método se centra en la *precisión* de la estimación, es decir, en el nivel de error de muestreo que consideramos aceptable. Como se vio en el Capítulo 6, el intervalo de confianza proporciona una indicación de la precisión de nuestra estimación de la muestra respecto al valor real de la población. Por tanto, esta estrategia se centra en la amplitud del intervalo de confianza que se obtendrá cuando se analicen los resultados del estudio. En el estudio de cáncer de mama, podemos estimar el tamaño de la muestra necesario para asegurar que el estudio pueda estimar la diferencia real en la supervivencia dentro de $\pm 2,5\%$ de su valor (es decir el intervalo de confianza se extenderá un $2,5\%$ a cada lado de la estimación de la muestra).

En este capítulo, empezamos considerando los cálculos del tamaño de la muestra basados en el poder y a continuación los cálculos basados en la precisión. El capítulo termina con una discusión de cómo aplicarlos a diseños de estudio más complejos y de otros aspectos prácticos que requieren ser considerados en cualquier estimación del tamaño de la muestra. Es importante destacar en esta fase que todo cálculo del tamaño de la muestra implica alguna conjetura ya que tiene que empezarse anticipando los resultados del estudio propuesto y, por tanto, debe considerarse que estos cálculos *sólo proporcionan una estimación cruda del tamaño de estudio requerido*. Además, como se verá más adelante en este capítulo, en toda situación práctica, aparte de las consideraciones estadísticas, también deben conside-

rarse otros aspectos (p.e. costes, disponibilidad de los sujetos elegibles, problemas logísticos).

15.2 Poder de un estudio

El *poder* de un estudio es la probabilidad de obtener un resultado 'estadísticamente significativo', es decir, un valor de P inferior a un nivel de 'significación' determinado preestablecido (normalmente 0,05), cuando la magnitud real del efecto es como la que se ha previsto. No obstante, como se ha comentado en el Apartado 6.3, existe una estrecha relación entre los valores de P y los intervalos de confianza. Por tanto, el poder también puede interpretarse como la *probabilidad de obtener una estimación cuyo intervalo de confianza no incluya el valor estipulado por la hipótesis nula*. La hipótesis nula establece que la exposición no tiene efecto sobre el resultado de interés correspondiente, con un valor de cero si el efecto de la exposición se mide en una escala absoluta (p.e. diferencia de riesgos o de tasas), o uno si se mide en una escala relativa (p.e. razón de riesgos o de tasas).

La Figura 15.1(a) muestra un diagrama que ilustra la relación entre el valor estipulado por la hipótesis nula, la estimación del efecto previsto y su intervalo de confianza esperado cuando la exposición está asociada con un aumento de la aparición del resultado de interés. Para que el estudio tenga un poder adecuado para detectar este efecto, el límite inferior del intervalo de confianza de la estimación del efecto previsto tiene que ser superior al valor estipulado por la hipótesis nula. Asimismo, cuando la exposición está asociada con una disminución del riesgo (es decir la exposición es protectora), el límite superior del intervalo de confianza debe ser inferior al valor estipulado por la hipótesis nula (Figura 15.1(b)). Por tanto, el poder del estudio para detectar un efecto 'estadísticamente significativo', si el efecto real es el previsto, es la probabilidad de que el límite inferior del intervalo de confianza sea mayor (o, si la exposición es protectora, el límite superior sea inferior) del valor estipulado por la hipótesis nula.

La posición del límite inferior (o del límite superior, si la exposición es protectora) del intervalo de confianza de la estimación del efecto previsto está determinada por la *amplitud del intervalo de confianza* ($\pm jEE$) (Figura 15.1), que a su vez depende del *tamaño del estudio* (cuanto mayor sea el estudio menor será el error estándar (EE) y, por tanto, más estrecho será el intervalo de confianza) y del *nivel de significación (confianza) elegido* (j). Para un intervalo de confianza al 95%, j sería igual a 1,96 (Tabla 15.2), es decir, el intervalo de confianza abarcará $1,96EE$ a cada lado de la estimación de la muestra. Para un intervalo de confianza al 99%, j será $2,576$ y, por tanto, el intervalo de confianza será más amplio. Cuanto más amplio sea el intervalo de confianza menor será el poder de un estudio de un tamaño dado.

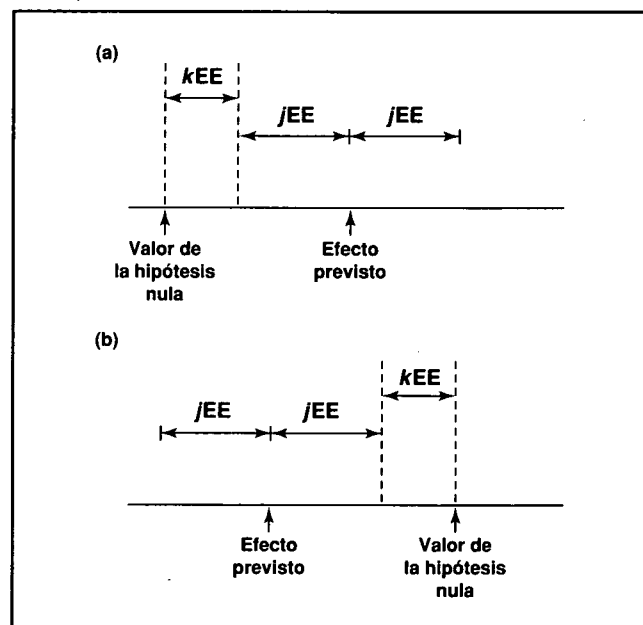


Figura 15.1. Diagrama que ilustra la relación entre el valor estipulado por la hipótesis nula, el efecto previsto y su intervalo de confianza cuando la exposición está asociada con un aumento (a) o una disminución (b), en la aparición del resultado de interés (adaptado de Clayton y Hills, 1993).

Nivel de significación	j
0,10	1,645
0,05	1,960
0,01	2,576
Poder	k
0,95	1,645
0,90	1,282
0,75	0,674
0,50	0,0
<0,50	<0

Tabla 15.2.
Valores de k y j para niveles de significación y poder diferentes.

Supongamos que el estudio se repite varias veces. Las estimaciones del efecto obtenidas cada vez y sus intervalos de confianza serían diferentes debido a la variación de muestreo. Si se representaran gráficamente las estimaciones del efecto obtenidas cada vez, obtendríamos una distribución normal de muestreo con un error estándar EE. Asimismo, si se representaran gráficamente los límites inferiores (o límites superiores, si la exposición es protectora) de cada uno de los intervalos de confianza, obtendríamos una distribución normal, con el mismo error estándar EE, centrado alrededor del valor previsto del límite inferior (o superior). El poder de un estudio es la probabilidad de que el límite inferior del intervalo de confianza se encuentre por encima del valor estipulado por la hipótesis nula (o el límite superior se encuentre por debajo, si la exposición es protectora). Esta probabilidad depende del *número de errores estándar* (k) entre la hipótesis nula y la posición prevista del límite inferior (o límite superior, si la exposición es protectora) del intervalo de confianza del efecto previsto estimado. Puede demostrarse matemáticamente que si k es igual a 1,645 el estudio tendrá un poder del 95% (Tabla 15.2). En otras palabras, si el estudio se realizara repetidamente, sería de esperar que sólo 5 de cada 100 intervalos de confianza al 95% resultantes incluyeran el valor estipulado por la hipótesis nula, si la magnitud real del efecto es la prevista. Cuando la posición prevista del límite de confianza inferior (o superior) se encuentra exactamente en la hipótesis nula, de modo que $k=0$, el poder es 0,50 y existe la misma probabilidad de obtener un resultado significativo que de no obtenerlo. Si $k<0$, el poder será menor del 50%. En general, un poder inferior al 80% se considera inaceptablemente bajo.

Por tanto, para resumir, el poder de un estudio depende de:

1. La *magnitud del efecto previsto* (es decir, la distancia entre el valor de la hipótesis nula y el efecto previsto). Cuanto mayor sea el efecto, mayor será el poder de detectarlo como 'estadísticamente significativo' para un estudio de un tamaño dado.
2. La *amplitud del intervalo de confianza* (jEE), que determina la posición del límite inferior (o del límite superior, si la exposición es protectora). Cuanto más amplio sea el intervalo de confianza menor será el poder de un estudio de un tamaño dado. Esto a su vez depende de:
 - (a) El tamaño del estudio. Cuanto mayor sea el estudio menor será el error estándar (EE) y, por tanto, más estrecho será el intervalo de confianza.
 - (b) El nivel de significación (confianza) elegido (j). Por ejemplo, un intervalo de confianza al 95% ($j=1,96$) será más estrecho que un intervalo de confianza al 99% ($j=2,576$) para una muestra de un tamaño dado.

Es útil construir curvas de poder para mostrar como varía el poder con el tamaño del estudio para diferentes niveles de significación y diferentes magnitudes del efecto previsto. La Figura 15.2 muestra algunos ejemplos de estas curvas para el estudio de cáncer de mama descrito en el Ejemplo 15.1.

En la mayoría de las situaciones reales, los investigadores tienen una idea muy buena del número de sujetos elegibles que podrán reclutar en su estudio. Este número normalmente se determina por la disponibilidad de sujetos elegibles, logística de reclutamiento, costes, etc. En estas circunstancias la cuestión importante no es «¿Qué tamaño debe tener el estudio?» sino «¿Es el número de sujetos disponible suficiente para proporcionar una respuesta clara a los objetivos del estudio?». Para responder a esta última pregunta, es necesario estimar el poder del estudio con el número de sujetos propuesto. Si estos cálculos muestran que el poder será demasiado bajo, será necesario estimar en cuánto es necesario aumentar el tamaño de la muestra para asegurar que el estudio alcance el poder deseado. Éste es el método propuesto por Clayton y Hills (1993), que seguiremos en este apartado.

15.2.1 Comparación de dos proporciones (prevalencias o riesgos)

Para calcular el poder de nuestro estudio de cáncer de mama (Ejemplo 15.1) para detectar una diferencia del 10% en la supervivencia entre los dos tratamientos, es necesario calcular en primer lugar el EE de la diferencia entre las dos proporciones de mujeres que permanecieron vivas al final del primer año. Como se vio en el Apartado A6.1.2, el EE de la diferencia entre dos proporciones puede estimarse, aproximadamente, como

$$EE(p_1 - p_0) = \sqrt{\frac{p_1^2(1-p_1)}{a} + \frac{p_0^2(1-p_0)}{b}}$$

Por tanto, en nuestro ejemplo

$$EE = \sqrt{\frac{0,80^2(1-0,80)}{80} + \frac{0,70^2(1-0,70)}{70}} = 0,061$$

La Figura 15.1(a) muestra que la distancia entre el efecto previsto y el valor estipulado por la hipótesis nula (en nuestro ejemplo, 0,10 - 0) es la suma de dos componentes, una procedente de la amplitud del intervalo de confianza (jEE) y otra de la distancia entre la posición prevista del límite inferior del intervalo de confianza y la hipótesis nula (kEE). Por consiguiente,

$$0,10 = jEE + kEE$$

Si el nivel de significación se establece en 0,05 ($j=1,96$) y $EE=0,061$, k será igual a

$$0,10 = 1,96 \times 0,061 + k \times 0,061$$

$$k = -0,32$$

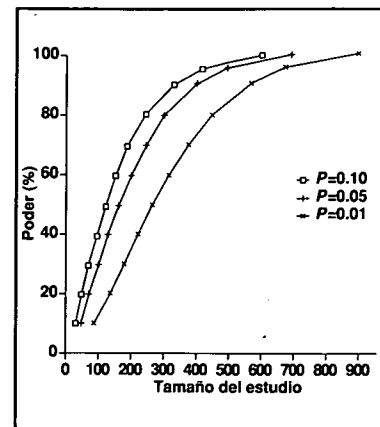


Figura 15.2.

Poder para detectar una diferencia del 10% en la supervivencia a un año cuando la supervivencia basal es del 70% (equivalente a un aumento del 14% en la supervivencia) para diferentes tamaños de muestra y niveles de significación.

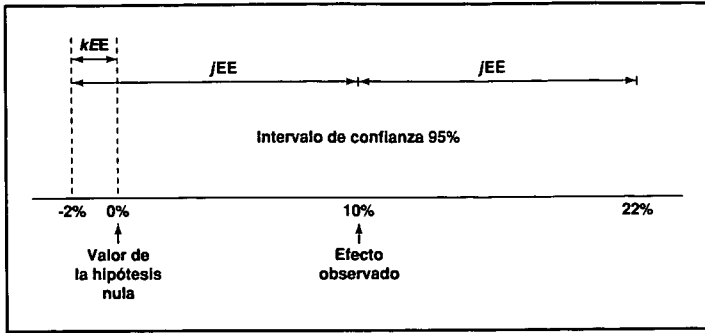


Figura 15.3. Cálculo del poder para el estudio clínico de cáncer de mama presentado en el Ejemplo 15.1: efecto observado = 10% (=80% - 70%) de diferencia en la supervivencia; tamaño de la muestra = 100 en cada grupo; nivel de significación = 0,05; poder < 0,50.

Este valor de k corresponde a un poder inferior al 50% (Tabla 15.2). Por tanto, la probabilidad de que el estudio pueda detectar una diferencia del 10% estadísticamente significativa en la supervivencia entre los dos tratamientos, incluso aunque esta diferencia exista realmente, es inferior al 50% porque el intervalo de confianza es demasiado amplio debido a la amplitud del EE. Esto se ilustra en la Figura 15.3.

Si suponemos que la diferencia real en la supervivencia es del 10% y fijamos el nivel de significación a 0,05 ($j=1,96$), podemos calcular el valor del error estándar que será necesario para asegurar que el poder del estudio sea 0,95 ($k=1,645$)

$$1,96 \times EE + 1,645 \times EE = 0,10$$

$$EE \text{ requerido} = 0,10 / (1,960 + 1,645) = 0,028$$

Este valor del EE es mucho menor que el valor obtenido con 100 sujetos en cada grupo ($EE = 0,061$). Para reducir el EE al valor requerido, es necesario aumentar el tamaño del estudio. Clayton y Hills (1993) proporcionaron la fórmula siguiente para calcular un factor por el cual debe aumentarse el tamaño del estudio para conseguir el poder especificado,

$$(\text{Valor actual de EE} / \text{Valor requerido de EE})^2$$

Así, en nuestro estudio

$$\text{Valor actual de EE} = 0,061$$

$$\text{Valor requerido de EE} = 0,028$$

$$\text{Factor de escalado} = 4,8$$

El tamaño inicial de la muestra era de 100 en cada grupo. Para asegurar que el estudio tenga un poder del 95% para detectar una diferencia del 10% en la supervivencia con un nivel de significación del 5%, es necesario multiplicar este tamaño de la muestra original por 4,8. Por tanto, necesitamos incluir 480 sujetos *en cada grupo de tratamiento*.

Como se vio en el Apartado A6.1.3, el test de significación es el mismo independientemente del tipo de medida del efecto (razón o diferencia) utilizada para comparar los dos grupos. Los cálculos del poder se basan en el test de significación que se realizará al final del estudio y, como consecuencia lógica, *se obtendrán unas estimaciones similares del tamaño de la muestra tanto si los cálculos se basan en la razón de las dos proporciones (80%/70%) como si se basan en su diferencia (80%-70%)*. El método sería similar al seguido anteriormente a excepción de que los cálculos se basarían en el EE de la razón de las dos proporciones en lugar del EE de su diferen-

cia. Debido a que el intervalo de confianza alrededor de una razón de proporciones es asimétrico, es decir, sólo puede tomar valores de cero a infinito (Apartado A6.1.3.2), primero es necesario convertir la razón de supervivencia estimada en su logaritmo natural (indicado por \ln)

$$\ln(0,80/0,70) = \ln(1,14) = 0,134$$

Ahora podemos calcular un EE 'aproximado' del logaritmo de la razón de dos proporciones (R) utilizando la fórmula indicada en el Apartado A6.1.2

$$\begin{aligned} EE(\ln R) &= \sqrt{(1/a + 1/b - 1/m_1 - 1/m_0)} \\ &= \sqrt{(1/80 + 1/70 - 1/100 - 1/100)} = 0,082 \end{aligned}$$

La distancia entre el logaritmo del efecto previsto y el logaritmo del valor de la hipótesis nula de 1 es igual a

$$\ln(1,14) - \ln(1) = 0,134 - 0 = 0,134$$

Por tanto,

$$0,134 = j \times EE + k \times EE$$

El valor del EE requerido para asegurar que el estudio tenga un poder del 95% para detectar una razón de riesgos de 1,14 al nivel de significación del 5% será

$$0,134 = 1,96 \times EE + 1,645 \times EE$$

$$EE \text{ requerido} = 0,134 / (1,960 + 1,645) = 0,037$$

Por tanto,

$$\text{Factor de escalado} = (0,082 / 0,037)^2 = 4,8$$

Este valor es exactamente el mismo que obtuvimos anteriormente cuando el cálculo se basó en la diferencia entre las dos proporciones en lugar de en su razón. Por tanto, el tamaño de la muestra necesario para detectar un aumento del 10% en la supervivencia (del 70% al 80%) es equivalente al tamaño de la muestra necesario para detectar una razón de riesgos de 80% / 70% = 1,14.

15.2.2 Comparación de dos tasas

Cálculos similares del tamaño de la muestra pueden realizarse para estudios de intervención y de cohortes en los cuales la razón de tasas (o la diferencia de tasas) es la medida apropiada del efecto.

Ejemplo 15.2. Supongamos que queremos realizar un estudio de cohorte para evaluar el efecto de una exposición ocupacional sobre la incidencia de cáncer de pulmón. Pretendemos incluir en el estudio 40.000 hombres de mediana edad y seguirlos durante cinco años. Se sabe que un 40% de los trabajadores están expuestos al riesgo y esperamos que la tasa en los expuestos sea dos veces la de los no expuestos (es decir, razón de tasas prevista = 2,0). Se estima que la tasa de incidencia de cáncer de pulmón en los no expuestos es de 50 casos por 100.000 personas-años. Los resultados de este estudio de cohorte se presentarán como en la Tabla 15.3.

Tabla 15.3.
Resultados de un estudio de cohorte hipotético: tabla simulada

	Exposición		Total
	Sí	No	
Nº de casos	a	b	n
Personas-años a riesgo	y_1	y_0	y
Tasa por 100.000 p-as	r_1	r_0	r

En el Ejemplo 15.2 se pueden predecir los valores de a y b utilizando el total de personas-tiempo (y) de observación en el estudio y la tasa de incidencia (r_0) de cáncer de pulmón estimada en los no expuestos. Debido a que el cáncer de pulmón es una enfermedad rara, podemos estimar el total de personas-tiempo de observación como:

$$y = 40.000 \times 5 \text{ años} = 200.000 \text{ p-as}$$

suponiendo que no hay pérdidas del seguimiento ni otras causas de muerte que compitan. Si el 40% de la cohorte está expuesta al factor de riesgo en estudio, entonces 0,60 no están expuestos. Por tanto,

$$y_1 = 200.000 \times 0,40 = 80.000 \text{ p-as}$$

$$y_0 = 200.000 \times 0,60 = 120.000 \text{ p-as}$$

Debido a que la razón de tasas (RT) prevista es igual a 2,0, los números de casos de cáncer de pulmón esperados entre trabajadores expuestos y no expuestos son

$$a = 80.000 \text{ p-as} \times (2 \times 50 \text{ por } 100.000 \text{ p-as}) = 80$$

$$b = 120.000 \text{ p-as} \times 50 \text{ por } 100.000 \text{ p-as} = 60$$

Ahora podemos completar la Tabla 15.3 con los resultados que prevemos obtener de este estudio de cohortes si nuestras suposiciones son correctas (Tabla 15.4).

	Exposición		Total
	Si	No	
Nº de casos	80 (<i>a</i>)	60 (<i>b</i>)	140 (<i>n</i>)
Personas-años a riesgo	80.000 (<i>y</i> ₁)	120.000(<i>y</i> ₀)	200.000 (<i>y</i>)
Tasa por 100.000 p-as	100 (<i>r</i> ₁)	50 (<i>r</i> ₀)	70 (<i>r</i>)

Tabla 15.4.
Resultados previstos del estudio de cohorte propuesto en el Ejemplo 15.2.

Como se muestra en el Apartado A6.1.2, un EE 'aproximado' del logaritmo de una razón de tasas puede calcularse como

$$EE(\ln RT) = \sqrt{(1/a + 1/b)}$$

Por tanto, en nuestro ejemplo

$$EE(\ln RT) = \sqrt{(1/80 + 1/60)} = 0,171$$

El número de EE entre el logaritmo de la razón de tasas prevista ($\ln(2,0) = 0,693$) y el logaritmo del valor de la hipótesis nula de ($\ln(1) = 0$) es

$$0,693 - 0 = j \times 0,171 + k \times 0,171$$

$$j + k = 0,693/0,171 = 4,05$$

Para un nivel de significación de 0,05, *j* sería igual a 1,96 y, por consiguiente,

$$k = 4,05 - 1,96 = 2,09$$

Este valor de *k* se corresponde con un poder superior a 0,95 (Tabla 15.2). Por tanto, la probabilidad de que este estudio de cohorte sea capaz de detectar una razón de tasas real de 2,0 (para el nivel de significación de 0,05) es mayor del 95%. Estimaciones similares del tamaño de la muestra se obtendrían si los cálculos se basaran en la diferencia de tasas prevista de 50 por 100.000 personas-años (100 por 100.000 personas-años - 50 por 100.000 personas-años).

Si con el número de sujetos propuesto, el poder del estudio fuera demasiado bajo, podríamos calcular en cuánto deberíamos aumentar el tamaño de la muestra para conseguir el nivel de poder deseable utilizando el procedimiento descrito anteriormente para comparar dos proporciones.

15.2.3 Comparación de dos odds

En estudios caso-control puede utilizarse un método similar. Para estimar el poder del estudio caso-control del Ejemplo 15.3 es necesario calcular los valores de *a*, *b*, *c*, *d*. Debido a que se supone que los controles son representativos de la población de la que proceden los casos, sería de esperar

Ejemplo 15.3. Supongamos que queremos realizar un estudio de 200 casos y 200 controles para detectar una odds ratio de 0,5 para una exposición determinada. Se sabe que la prevalencia de esta exposición en la población de la que proceden los casos es del 33,5%. Los resultados que se obtendrán a partir de este estudio caso-control se presentarán como en la Tabla 15.5.

Tabla 15.5.
Resultados de un estudio caso-control hipotético.

	Exposición		Total
	Sí	No	
Casos	a	b	n_1
Controles	c	d	n_0
Total	m_1	m_0	N

que un 33,5% de ellos estuviera expuesto y el resto no expuesto al factor en investigación. Por tanto

$$c = 200 \times 0,335 = 67$$

$$d = 200 - 67 = 133$$

Ahora podemos calcular la odds de la exposición entre los controles como

$$c/d = 67 / 133 = 0,5$$

Debido a que prevemos una odds ratio igual a 0,5, podemos calcular la odds de la exposición entre los casos como

$$a/b = 0,5 \times 0,5 = 0,25$$

Por tanto,

$$a = 0,25 \times b$$

Como $a + b = 200$, entonces

$$b = 160$$

$$a = 200 - 160 = 40$$

Ahora podemos completar la Tabla 15.5 con los valores de a , b , c y d que esperamos observar en el estudio caso-control propuesto (Tabla 15.6). Se puede calcular un EE 'aproximado' del logaritmo de la odds ratio prevista utilizando la fórmula indicada en el Apartado A6.1.2.

	Exposición		Total
	Sí	No	
Casos	40 (a)	160 (b)	200 (n ₁)
Controles	67 (c)	133 (d)	200 (n ₀)
Total	107 (m₁)	293 (m₀)	400 (N)

Tabla 15.6.
Resultados previstos del estudio caso-control hipotético descrito en el Ejemplo 15.3.

$$EE = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

Por tanto, en nuestro estudio caso-control

$$EE = \sqrt{(1/40 + 1/160 + 1/67 + 1/133)} = 0,232$$

Para calcular el poder del estudio para un nivel de significación de 0,05 ($j = 1,960$), primero es necesario calcular $\ln(\text{OR}) = \ln(0,5) = -0,693$. Éste es un valor negativo, pero como estamos interesados en la distancia absoluta entre el valor previsto y la hipótesis nula, puede ignorarse el signo negativo y proceder como habitualmente

$$\begin{aligned} 0,693 &= 1,96 \times EE + k \times EE \\ 0,693 - 1,96 \times 0,23 &= k \times 0,23 \\ 0,238 &= k \times 0,23 \\ k &= 1,03 \end{aligned}$$

Un valor k de 1,03 corresponde a un poder cercano a 0,85 (Tabla 15.2). Si queremos aumentar el poder a 0,95, tiene que aumentarse el tamaño del estudio.

$$0,693 = 1,96 \times EE + 1,645 \times EE$$

$$EE \text{ requerido} = 0,693 / (1,96 + 1,645) = 0,192$$

$$\text{Factor de escalado} = (0,232/0,192)^2 = 1,46$$

Por tanto, si queremos aumentar el poder del estudio de cerca de 0,85 hasta 0,95, el tamaño propuesto de la muestra debe multiplicarse por un factor de 1,46, es decir, de 200 a 292.

15.3 Cálculos del tamaño de la muestra basados en la precisión

Los cálculos del tamaño de la muestra comentados en el apartado anterior se centraban en el test de significación estadística que se realizará al final del estudio. La principal limitación de este método es que puede producir estimaciones muy imprecisas de la asociación con el resultado

Ejemplo 15.4. Volvamos a considerar el estudio de cáncer de mama (Ejemplo 15.1). Para que el estudio tenga un poder del 95% para detectar una diferencia de 10% en la supervivencia, al nivel de significación del 5%, el tamaño de la muestra debe aumentarse de 100 a 480 sujetos en cada uno de los dos grupos de tratamiento (Apartado 15.2.1). Con este nuevo tamaño de muestra, prevemos los siguientes resultados (Tabla 15.7 y Figura 15.4).

Tabla 15.7.

Resultados previstos de un estudio hipotético para evaluar el valor de un tratamiento nuevo sobre la supervivencia de cáncer de mama a un año. Efecto previsto = 10% (= 80%-70%) de diferencia en la supervivencia; nivel de significación = 0,05; poder = 0,95; tamaño de la muestra = 480 mujeres en cada grupo de tratamiento.

		Tratamiento nuevo	Tratamiento estándar	Total
Viva 1 año después del diagnóstico	Sí	384 (a)	336 (b)	720 (n_1)
	No	96 (c)	144 (d)	240 (n_0)
	Total	480 (m_1)	480 (m_0)	960 (N)

$p_1 = 384/480 = 80\%$

$p_0 = 336/480 = 70\%$

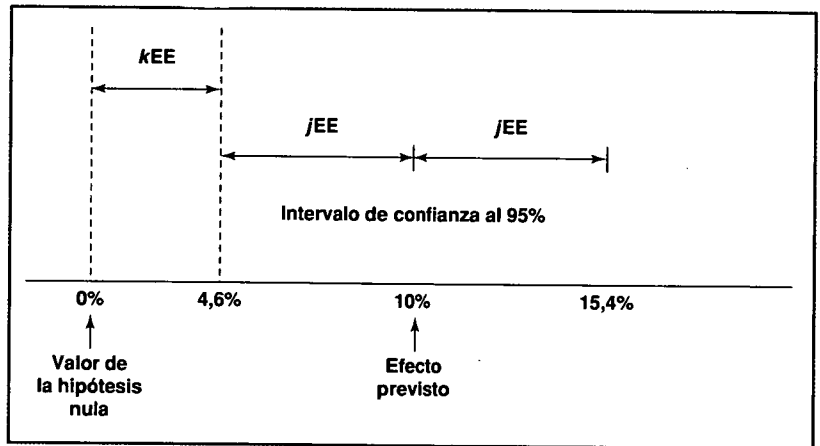
Diferencia de riesgo = $p_1 - p_0 = 10\%$

Intervalo de confianza al 95% = +4,6% a + 15,4%

$\chi^2 = 12,79$ con 1 g.l.; $P < 0,001$

Figura 15.4.

Efecto previsto y su intervalo de confianza al 95% para el estudio hipotético de cáncer de mama descrito en el Ejemplo 15.4: efecto previsto = 10% (= 80%-70%) de diferencia en la supervivencia; nivel de significación = 0,05; poder = 0,95; tamaño de la muestra = 480 mujeres en cada grupo de tratamiento.



de interés, es decir, aunque el intervalo de confianza no incluya el valor de la hipótesis nula, puede resultar demasiado amplio para ser informativo.

En el Ejemplo 15.4, aunque el intervalo de confianza previsto no incluye el valor de la hipótesis nula de ninguna diferencia entre los dos tratamientos, su amplitud es compatible con una mejoría en la supervivencia a un año, que oscila del 4,6% al 15,4% (Tabla 15.7; Figura 15.4). Ésta no es una estimación muy precisa. En realidad, este intervalo es más amplio que el 10% de diferencia prevista. Podemos considerar como más aceptable una amplitud de $\pm 2,5\%$ a cada lado de la estimación muestral del 10%, es decir, un intervalo de confianza de la diferencia en la proporción de muje-

res todavía vivas al final del primer año que oscile entre el 7,5% y el 12,5%.

Los cálculos del tamaño de la muestra basados en el poder pueden ser apropiados para exposiciones nuevas cuando muchas veces no se sabe si existirá un impacto sobre los resultados de interés. No obstante, si otros estudios ya han demostrado que la exposición está asociada con un aumento o una disminución del riesgo, no tiene mucho sentido contrastar la hipótesis nula y el objetivo debe ser estimar la magnitud del efecto con la mayor precisión posible. En estas situaciones, es más apropiado elegir el tamaño de la muestra que proporcionará un intervalo de confianza de una amplitud predefinida.

Ejemplo 15.5. *Supongamos que queremos realizar un estudio transversal en una zona determinada para estimar la prevalencia del uso actual de anticonceptivos orales en mujeres de 20-44 años. Pensamos tomar una muestra aleatoria de la población de todas las mujeres de 20-44 años que viven en esta área. Queremos calcular el tamaño de la muestra necesario para asegurar que el estudio pueda estimar la prevalencia real del uso actual de anticonceptivos orales en el área en estudio dentro de un 5% de su valor (e.d., el intervalo de confianza que obtendremos cuando los resultados del estudio se analicen se extenderá un 5% a cada lado de la estimación de la muestra).*

15.3.1 Estimación de una proporción cruda única (prevalencia o riesgo)

Podemos estimar el tamaño de la muestra necesario para asegurar que el intervalo de confianza para una proporción única (prevalencia o riesgo) sea de una anchura determinada.

En el Ejemplo 15.5, pensamos tomar una muestra de n mujeres de 20-44 años. Si entre ellas, a son usuarias actuales, estimamos la prevalencia del uso de anticonceptivos orales como

$$p = a/n$$

Como se vio en el Apartado 6.1, esta estimación está sometida a error de muestreo, pero el intervalo de confianza al 95% incluirá la prevalencia real del uso de anticonceptivos orales en la población con una probabilidad del 95%.

$$EE(p) = \sqrt{\frac{p^2(1-p)}{a}}$$

Supongamos que queremos que nuestro intervalo de confianza alrededor de la estimación de la muestra tenga una amplitud determinada ($\pm w$). El valor de w depende del error estándar EE y del nivel de significación (j) elegido.

$$w = j \times EE$$

Para un intervalo de confianza del 95%, $j = 1,96$, es decir, el intervalo se extiende 1,96 errores estándar a cada lado de la estimación ($w = 1,96 \times EE$).

Por consiguiente, podemos estimar la prevalencia del uso de anticonceptivos orales (p) con un grado de precisión predefinido eligiendo un tamaño de la muestra (n) apropiado. En primer lugar, tenemos que prever el valor de p . Supongamos que las estadísticas sobre ventas de anticonceptivos orales indican que la prevalencia del uso es probable que sea de alrededor del 50% y queremos estimarla dentro del $\pm 5\%$. Por tanto

$$p = 0,50$$

$$w = j \times EE = 0,05$$

Si elegimos un nivel de confianza del 95%, entonces

$$1,96 \times EE = 0,05$$

$$EE(p) = 0,05 / 1,96 = 0,0255$$

Como que $p = 0,5$, podemos estimar a a partir de la fórmula para el EE,

$$EE(p) = \sqrt{\frac{p^2(1-p)}{a}} = 0,0255$$

$$\sqrt{(0,5^2 \times 0,5) / a} = 0,0255$$

$$0,125 / a = 0,0255^2$$

$$a = 192$$

Por último, podemos calcular el tamaño de la muestra (n) requerido como

$$p = a/n$$

$$0,5 = 192 / n$$

$$n = 384$$

Por tanto, necesitaremos incluir 384 mujeres en el estudio. Al planificar un estudio es una buena idea determinar los tamaños de muestra que serán necesarios para diferentes niveles de precisión (Tabla 15.8 y Figura 15.5).

Por tanto, para estimar una prevalencia real del uso de anticonceptivos orales del 50% dentro de $\pm 1\%$ (es decir, del 49% al 51%) necesitaríamos incluir en el estudio a 9.612 mujeres.

También es importante calcular el tamaño de la muestra requerido para diferentes valores de la prevalencia prevista p . Como puede observarse en la Tabla 15.9, el tamaño de la muestra necesario no varía mucho para valores de p entre 0,3 y 0,7, siendo máximo cuando $p = 0,50$. Por tanto, para asegurarnos, podemos establecer $p = 0,50$ y obtener el tamaño máximo de la muestra (n) requerido.

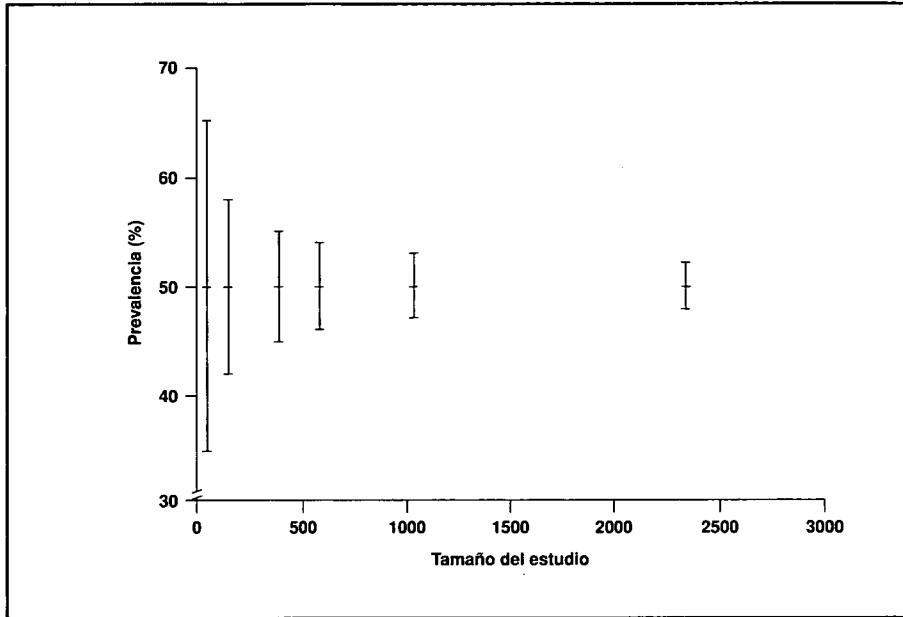


Tabla 15.8.

Tamaños de muestra necesarios para estimar una prevalencia real del 0,50 con intervalos de confianza al 95% de diferentes amplitudes ($\pm w$).

Amplitud ($\pm w$)	Tamaño de la muestra (n)
0,01	9.612
0,02	2.403
0,03	1.068
0,04	600
0,05	384
0,06	266
0,07	196
0,08	150
0,09	118
0,10	96
0,15	43

Figura 15.5.

Intervalos de confianza al 95% previstos para una prevalencia real de 0,50 según diferentes tamaños de muestra.

15.3.2 Estimación de una tasa cruda única

Un método similar puede utilizarse para estimar el tamaño de la muestra necesario para asegurar que el intervalo de confianza de una tasa individual sea de una amplitud predeterminada.

Ejemplo 15.6. Supongamos que queremos determinar la tasa de incidencia de una enfermedad particular en una población determinada. Basándonos en los datos de estudios previos, prevemos una tasa de 50 por 10.000 p-as. Queremos estimar la tasa de incidencia dentro de unos márgenes de ± 5 por 10.000 personas-año.

En el Ejemplo 15.6 pensamos tomar una muestra aleatoria de individuos de la población en estudio.

Por tanto, para un nivel de confianza del 95% ($j = 1,96$)

$$w = 1,96 \times EE$$

$$5 \text{ por } 10.000 \text{ p-as} = 1,96 \times EE(r)$$

$$EE(r) = 2,55 \text{ por } 10.000 \text{ p-as}$$

Prevalencia (p)	Tamaño de la muestra (n)
0,10	138
0,20	245
0,30	322
0,40	367
0,50	384
0,60	367
0,70	322
0,80	245
0,90	138

Tabla 15.9.

Tamaño de la muestra requerido para asegurar que el intervalo de confianza al 95% para diferentes niveles de prevalencia (p) se amplíe $w = 0,05$ a cada lado de la estimación de la muestra.

El error estándar 'aproximado' de una tasa puede calcularse como se indica en el Apartado A.6.1.1:

$$EE(r) = \frac{r}{\sqrt{a}}$$

donde r es la tasa estimada y a es el número de casos que se produjeron durante el periodo de observación. Por tanto, en nuestro ejemplo

$$\frac{50 \text{ por } 10.000 \text{ p-as}}{\sqrt{a}} = 2,55 \text{ por } 10.000 \text{ p-as}$$

$$a = 384$$

Ahora podemos calcular el número de personas-tiempo a riesgo (y) requerido para originar 384 casos

$$\begin{aligned} r &= a / y \\ 50 \text{ por } 10.000 \text{ p-as} &= 384 / y \\ y &= 76.800 \text{ p-as} \end{aligned}$$

Este conjunto de personas-años a riesgo total puede conseguirse siguiendo 76.800 individuos durante un año o 38.400 durante dos años, etc.

15.3.3 Estimación de una diferencia o razón de dos proporciones (riesgos o prevalencias)

Volvamos a considerar el estudio de cáncer de mama (Ejemplos 15.1 y 15.4). Supongamos que queremos asegurar que la amplitud del intervalo de confianza para la diferencia en las proporciones sea igual a $\pm w = 2 \times 2,5\%$. Por tanto,

$$w = 1,96 \times EE (p_1 - p_0) = 0,025$$

El EE requerido debe ser

$$0,025 / 1,96 = 0,0128$$

y

$$EE (p_1 - p_0) = \sqrt{\frac{p_1^2(1 - p_1)}{a} + \frac{p_0^2(1 - p_0)}{b}} = 0,0128$$

Como $a = 0,80m_1$, $b = 0,70m_0$ y $m_0 = m_1$, entonces

$$\begin{aligned} m_0 &= m_1 = a / 0,80 \\ b &= 0,70 \times (a/0,80) = 0,875 a \end{aligned}$$

La fórmula anterior para el error estándar puede escribirse como

$$EE = \sqrt{\frac{0,80^2(1-0,80)}{a} + \frac{0,70^2(1-0,70)}{0,875a}} = 0,0128$$

$$(0,0128)^2 = 0,128 / a + 0,147 / 0,875a$$

$$0,000164 = ((0,128 \times 0,875) + 0,147) / 0,875a$$

$$0,875a = 0,259 / 0,000164 = 1.579,3$$

$$a = 1.805$$

$$m_1 = m_0 = 1.805 / 0,80 = 2.256$$

$$b = 2.256 \times 0,70 = 1.579$$

Por tanto, para obtener un intervalo de confianza del 95% para la diferencia de supervivencia entre los dos tratamientos con una amplitud del 2,5% a cada lado del efecto previsto del 10% (es decir, del 7,5% al 12,5%), es necesario incluir 2.256 sujetos en cada grupo de tratamiento.

Asimismo, podemos calcular el tamaño de la muestra requerido para estimar una razón de dos proporciones con un nivel de precisión predefinido. El método sería similar al ilustrado anteriormente, excepto en que los cálculos se basarían en la fórmula para el EE de una razón de proporciones (véase Anexo A6.1.2).

15.3.4 Estimación de una diferencia de tasas o una razón de tasas

En estudios en los que la medida adecuada del efecto es la razón de tasas puede utilizarse un método similar.

Ejemplo 15.7. La incidencia de cáncer de estómago en hombres de 50-59 de una población determinada es 65 casos por 100.000 personas-año. Supongamos que estamos planeando realizar un estudio en esta población para evaluar si una intervención determinada puede reducir la tasa de cáncer de estómago en hombres de este grupo de edad. Los sujetos elegibles se distribuirán aleatoriamente para recibir la intervención o el placebo en una relación de 1:1. Se espera que la tasa entre los que reciben la intervención sea del 60% de la tasa en aquellos a los que se administra placebo (es decir, una reducción del 40% del riesgo). Queremos que el intervalo de confianza para la estimación de la razón de tasas tenga una amplitud (en la escala logarítmica) de 1,30 a cada lado de la estimación de la muestra (es decir, de 0,46 (= 0,60/1,30) a 0,78 (= 0,60 × 1,30)). Los resultados de este estudio se presentarán como en la Tabla 15.3.

Un intervalo de confianza al 95% 'aproximado' del logaritmo de una razón de tasas (r_1/r_0) puede estimarse como

$$\text{Intervalo de confianza al 95\%} = \ln RT \pm 1,96 \times EE (\ln RT)$$

En el Ejemplo 15.7,

$$w = \ln(1,30) = 1,96 \times EE(\ln RT)$$

$$EE(\ln RT) = 0,262 / 1,96 = 0,134$$

Un error estándar 'aproximado' del logaritmo de una razón de tasas (RT) estimada puede obtenerse como sigue

$$EE(\ln RT) = \sqrt{(1/a + 1/b)}$$

En nuestro ejemplo,

$$0,134 = \sqrt{(1/a + 1/b)}$$

Debido a que debe asignarse el mismo número de sujetos a cada rama del estudio y prevemos que la razón de tasas sea igual a 0,6, entonces

$$a = 0,6 \times b$$

$$0,134 = \sqrt{(1/0,6b + 1/b)} = \sqrt{(1,6/0,6b)}$$

$$(0,134)^2 = 1,6/0,6b$$

$$0,6b = 1,6/0,018 = 88,89$$

$$b = 88,89/0,6 = 148$$

$$a = 0,6 \times 148 = 89$$

La tasa de incidencia de cáncer de estómago para los que se administró placebo se cree que es de 65 por 100.000 personas-año. Por tanto, necesitamos 227.692 personas-año (= 148/0,00065 personas-año) de seguimiento en cada rama del estudio para poder acumular 148 casos de cáncer de estómago entre los que recibieron la intervención y 89 entre los que se administró placebo. Esto garantizará que el resultado del ensayo tenga el nivel deseado de precisión. Esto puede conseguirse siguiendo 227.692 hombres durante un año, 113.846 durante dos años y así sucesivamente.

Un método similar puede utilizarse para calcular estimaciones del tamaño de la muestra basadas en la precisión en situaciones en las que la medida apropiada del efecto es la diferencia de tasas en lugar de la razón de tasas (véase Anexo A6.1.2).

15.3.5 Estimación de la *odds ratio*

Consideremos el ejemplo siguiente.

Ejemplo 15.8. Supongamos que queremos realizar un estudio de casos y controles para detectar una *odds ratio* de 2,0 para una exposición presente en un 33,5% de la población de la que proceden los casos. En el estudio se incluirá un número similar de casos y controles. Queremos asegurar que el intervalo de confianza del 95% para la *odds ratio* tenga una amplitud (en la escala logarítmica) de 1,25 a cada lado de la estimación de la muestra (es decir, de 1,6 (= 2,0/1,25) a 2,5 (= 2,0 × 1,25)). Los resultados de este estudio caso-control se presentarán como en la Tabla 15.5.

El intervalo de confianza al 95% del logaritmo de una *odds ratio* puede estimarse como

$$\text{Intervalo de confianza al 95\%} = (\ln \text{OR}) \pm 1,96 \times \text{EE} (\ln \text{OR})$$

En nuestro ejemplo,

$$w = \ln(1,25) = 1,96 \times \text{EE} (\ln \text{OR})$$

$$\text{EE} (\ln \text{OR}) = 0,223 / 1,96 = 0,114$$

Un error estándar 'aproximado' del logaritmo de la *odds ratio* estimada puede obtenerse como

$$\text{EE} (\ln \text{OR}) = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

Se espera que la *odds* de exposición entre los casos sea el doble de la *odds* de exposición entre los controles

$$a/b = 2 \times (c/d)$$

y la prevalencia de exposición entre los controles se espera que sea del 33,5%

$$c = n_0 \times 0,335$$

$$d = n_0 \times 0,665$$

Por lo tanto,

$$a/b = 2 \times \frac{n_0 \times 0,335}{n_0 \times 0,665} = 1,008$$

$$a = 1,008 \times b$$

$$n_0 = n_1 = a + b = 1,008b + b = 2,008b$$

$$0,114 = \sqrt{\frac{1}{1,008b} + \frac{1}{b} + \frac{1}{2,008b \times 0,335} + \frac{1}{2,008b \times 0,665}}$$

$$0,114^2 = 1/b \times (1/1,008 + 1 + 1/0,673 + 1/1,335)$$

$$0,013 = 4,227 / b$$

$$b = 325$$

$$a = 328$$

$$c = (325 + 328) \times 0,335 = 219$$

$$d = (325 + 328) \times 0,665 = 434$$

$$n_0 = a + b = 325 + 328 = 653$$

$$n_1 = c + d = 219 + 434 = 653$$

Por tanto, es necesario incluir 653 casos y 653 controles en el estudio caso-control para obtener al final del estudio un resultado con el nivel de precisión especificado.

15.4 Otras consideraciones en relación al tamaño del estudio

Es necesario precisar que los métodos descritos en este capítulo sólo proporcionan una estimación aproximada del tamaño del estudio requerido, ya que se basan en conjeturas o estimaciones aproximadas de los parámetros, decisiones subjetivas sobre el tamaño de un efecto que querríamos detectar y el uso de fórmulas aproximadas. Por lo tanto, estos métodos sólo proporcionan una idea del tipo de tamaño necesario.

En la práctica, habrá restricciones en los recursos que pueden limitar el tamaño máximo posible de la muestra. Todos los recursos en términos de personal, vehículos, capacidad del laboratorio, tiempo y dinero pueden estar limitados y normalmente es necesario conseguir un equilibrio entre los resultados de los cálculos del tamaño del estudio y lo que realmente puede realizarse considerando los recursos. Es probable que intentar realizar un estudio cuyos costes estén por encima de los recursos disponibles resulte infructuoso, ya que puede afectarse la calidad de los datos y el estudio puede colapsarse antes de terminarlo, desperdiciando el tiempo y el dinero que ya se habían invertido en el estudio. Por otro lado, si los cálculos muestran que un estudio de un tamaño factible tendrá un poder y/o una precisión inaceptablemente bajos, puede ser recomendable incluir otros centros.

El cálculo del tamaño del estudio siempre debe efectuarse para varios escenarios diferentes, en lugar de uno sólo (p.e., niveles diferentes de poder/precisión y de estimaciones de tasas/proporciones/riesgos relativos), para proporcionar un perfil claro del objetivo del estudio. Un método útil en la decisión de la relación entre coste y poder es construir curvas de poder para una o dos de las variables clave del estudio, para mostrar como varía el poder con el tamaño del estudio para diferentes valores de la medida del efecto (como se muestra en la Figura 15.2).

15.4.1 Estudios diseñados para explorar múltiples relaciones exposición-resultado

Muchos estudios epidemiológicos están diseñados para observar varias relaciones exposición-resultado. Podemos utilizar los métodos descritos en este capítulo para calcular el tamaño del estudio necesario para poder detectar las relaciones exposición-resultado más importantes. En el caso ideal, elegiríamos el más grande como nuestro tamaño de estudio idóneo. Puede ser que el tamaño del estudio requerido para detectar una o más de las relaciones exposición-resultado en que estamos interesados esté claramente por encima de nuestros recursos disponibles. Por ejemplo, podemos estar interesados en examinar la relación entre el peso al nacer con dos variables: patologías benignas de mama y cáncer de mama. Los cálculos del tamaño del estudio pueden mostrar que nuestra muestra será lo suficientemente grande para asegurar que el estudio tenga el suficiente poder/precisión para examinar la relación entre el peso al nacer y la aparición de enfermedades de mama benignas, pero resultará demasiado pequeña para examinar la relación entre el peso al nacer y la incidencia de cáncer de mama. Si no podemos aumentar el tamaño de la muestra al nivel deseado, tendremos que limitar los objetivos del estudio y considerar la enfermedad benigna de mama como el único resultado de interés.

15.4.2 Rechazos y pérdidas en el seguimiento

Los rechazos y/o pérdidas en el seguimiento pueden producirse en la mayoría de los estudios epidemiológicos porque los sujetos se trasladen de la zona del estudio, fallezcan por alguna causa no relacionada con la enfermedad de interés, se nieguen a continuar con el estudio, etc. Estas pérdidas reducen el tamaño del estudio disponible para el análisis y, por tanto, disminuyen el poder o la precisión del estudio. Podemos compensar estas pérdidas aumentando el tamaño inicial del estudio. Por ejemplo, si los cálculos del tamaño del estudio indican que son necesarios 320 sujetos y se espera una pérdida del seguimiento del 20%, el tamaño del estudio debe aumentarse a aproximadamente 400 ($400 - 400 \times 0,20 = 320$). No obstante, cabe destacar que esta estrategia asegurará que el estudio tenga el poder o la precisión necesario pero no evitará la posibilidad de sesgo de selección, ya que los individuos que se niegan a participar o se pierden del seguimiento pueden diferir en aspectos importantes de los que permanecen en el estudio.

15.4.3 Grupos de tamaños diferentes

Existen situaciones en las que podemos querer estudiar grupos de diferente tamaño. Por ejemplo, cuando el número de casos y controles disponibles es grande y el coste de obtener información de los dos grupos es comparable, la relación óptima control-caso es 1:1. No obstante, si el número de casos disponible para el estudio es limitado, o cuando el coste de obtener información es mayor para los casos que para los controles, el número de controles por caso puede aumentarse para conseguir el poder o la

precisión necesarios. Por ejemplo, un estudio con 100 casos y 100 controles tiene el mismo poder que un estudio con 75 casos y 150 controles, o uno con 63 casos y 252 controles, o uno con 55 casos y 550 controles. Normalmente no se recomienda aumentar la relación control:caso por encima de 4:1 porque con cada control adicional por encima de este punto tan sólo se produce un ligero aumento en el poder estadístico. Este pequeño aumento en el poder generalmente incrementa de forma considerable las necesidades logísticas y los costes asociados al reclutamiento de un número de sujetos mucho mayor, a no ser que la inclusión de los controles se obtenga a un coste adicional muy bajo.

15.4.4 Confusión e interacción

Las estimaciones del tamaño de la muestra presentadas aquí se refieren a medidas *crudas*. No obstante, en la mayoría de los estudios, es esencial controlar las variables de confusión mediante el cálculo de medidas *ajustadas*. Estos ajustes normalmente dan lugar a una pérdida del poder o de la precisión del estudio. No obstante, esta pérdida sólo es sustancial cuando existe una variable de confusión muy fuerte (Smith y Day, 1984).

Por el contrario, tal y como se comentó en el capítulo anterior, los estudios diseñados con el objetivo de detectar interacciones requieren unos tamaños de la muestra mucho mayores que los diseñados para observar efectos simples (Smith y Day, 1984). Esto limita en la práctica el número de estudios que pueden realizarse de manera explícita para examinar interacciones.

15.4.5 Diseños de estudio especiales

Los cálculos del tamaño de la muestra para estudios caso-control presentados en este capítulo se refieren a estudios no apareados. Los métodos utilizados en el cálculo del tamaño de la muestra adecuado para estudios caso-control apareados se presentan en Breslow y Day (1980).

Los métodos presentados en este capítulo suponen que las unidades de estudio son los *individuos* en lugar de los *grupos*. Los cálculos serían diferentes si, por ejemplo, la unidad de estudio fueran comunidades (Smith y Morrow, 1996).

Cuadro 15.1. Puntos esenciales

- Los cálculos del tamaño de la muestra pueden realizarse para asegurar que:
 1. El estudio posee el *poder* suficiente, es decir, la capacidad de detectar un resultado estadísticamente significativo si la magnitud real del efecto es como la que se ha previsto. Por tanto, esta estrategia se centra en el test de significación que se realizará al final del estudio.
 2. Las estimaciones de la muestra son *precisas*, es decir, el nivel de error de muestreo es bajo. Esta estrategia se centra en la amplitud del intervalo de confianza.
- Los cálculos del tamaño de la muestra basados en el poder pueden ser apropiados para identificar exposiciones nuevas cuando no se sabe si tendrán algún efecto sobre el(los) resultado(s) de interés. No obstante, si ya se sabe que la exposición está asociada con el resultado, el objetivo del estudio debe ser cuantificar la magnitud del efecto con la mayor precisión posible en lugar de sólo contrastar la hipótesis nula de falta de efecto. En estas circunstancias, los cálculos del tamaño de la muestra deben basarse en la precisión.
- Los cálculos del tamaño de la muestra deben considerarse como indicadores aproximados del número de sujetos que es necesario reclutar. Otras consideraciones tales como la disponibilidad de los sujetos, recursos, costes, etc. deben tenerse en cuenta detenidamente.

Lecturas adicionales

- * Los cálculos del poder presentados en el Apartado 15.2 siguen la metodología propuesta por Clayton y Hills (1993).
- * El cálculo del tamaño de la muestra para diseños de estudios de cohorte, caso-control e intervención más complejos se presentan respectivamente en Breslow y Day (1987, 1980) y en Smith y Morrow (1996).

Capítulo 16

Prevención del cáncer

16.1 Introducción

Los capítulos anteriores de este libro se han centrado en los principios y métodos necesarios para estudiar los determinantes de enfermedad y sus efectos. El objetivo último de la epidemiología es proporcionar conocimientos que ayuden a la formulación de programas de salud pública destinados a prevenir el desarrollo de enfermedad en personas sanas.

Prácticamente todos los aspectos importantes en la prevención del cáncer están relacionados con la historia natural de la enfermedad 'cáncer', que puede resumirse como se muestra en la Figura 16.1. En esta figura, el punto A indica la aparición biológica de la enfermedad y el inicio de la fase preclínica. Éste puede ser el punto en el que tiene lugar una serie irreversible de acontecimientos (p.e. mutación celular). Como resultado de la progresión de la enfermedad, aparecen los síntomas y/o signos que llevan al paciente a la atención médica y diagnóstico en el punto C. Éste es el final de la fase preclínica, que es el periodo desde A a C y el inicio de la fase clínica de la historia natural. La enfermedad puede progresar entonces a curación (D_1), enfermedad permanente e incapacidad (D_2) o muerte (D_3). Obviamente, el tiempo desde los síntomas y/o signos iniciales hasta la curación, enfermedad permanente o muerte, puede reflejar los efectos de tratamientos administrados, así como las características subyacentes de la enfermedad no tratada.

En este esquema está implícita la noción de que una enfermedad evoluciona en el tiempo y que, a medida que esto se produce, los cambios patológicos pueden volverse irreversibles. El objetivo de la prevención es interrumpir esta progresión.

Existen varios niveles de prevención:

- * *Prevención primaria* es la prevención de la enfermedad mediante una reducción de la exposición de los individuos a los factores de riesgo o mediante un aumento de su resistencia a ellos, evitando así la aparición de A.
- * *Prevención secundaria* (aplicada durante la fase preclínica) es la detección y el tratamiento precoz de la enfermedad. Las actividades de cribaje son un componente importante de la prevención secundaria. En

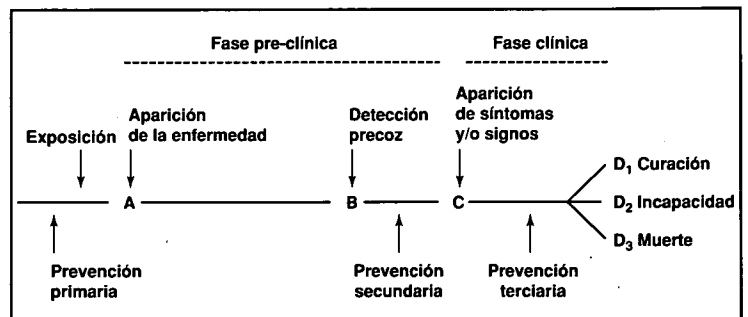


Figura 16.1. Historia natural de la enfermedad 'cáncer' y niveles de prevención.

el diagrama mostrado en la Figura 16.1, el punto B indica el momento en el tiempo en el que la enfermedad es detectable por una prueba de cribaje apropiada. Por ejemplo, el punto B puede referirse al momento en que una masa cancerosa alcanza el tamaño mínimo que puede observarse por rayos X. Por tanto, la distancia desde el punto B al C representa la '*fase preclínica detectable*'. Obviamente, la posición del punto B varía considerablemente de un individuo a otro. También depende de la técnica de cribaje utilizada.

- * *Prevención terciaria* (apropiada en la fase clínica) es el uso del tratamiento y programas de rehabilitación para mejorar el desenlace de la enfermedad entre los individuos afectados.

En este capítulo consideraremos en detalle cada uno de estos niveles de prevención.

16.2 Prevención primaria

El objetivo de la prevención primaria es limitar la incidencia de cáncer mediante el control de la exposición a factores de riesgo o el aumento de la resistencia de los individuos a los mismos (p.e. vacunación, quimioprolifaxis). Claramente, el primer paso es identificar las exposiciones pertinentes y evaluar su impacto sobre el riesgo de la enfermedad en la población.

16.2.1 ¿Qué importancia tiene una exposición en particular?

Medidas relativas y absolutas del efecto de la exposición

Gran parte de la epidemiología se ocupa de la identificación de los factores de riesgo para la enfermedad, problema sanitario o estado de salud. En la evaluación de la fuerza de la asociación entre una exposición y un resultado concretos, se calculan las medidas conocidas como *medidas relativas del efecto*. Como se indica en el Apartado 5.2.1, existen tres tipos de medidas relativas (razón de riesgos, razón de tasas y *odds ratio*) que con frecuencia se denominan en conjunto riesgo relativo.

Las medidas relativas del efecto responden a la pregunta: *¿Cuántas veces es más probable que los expuestos a un factor de riesgo putativo desarrollen el resultado de interés respecto a los no expuestos, asumiendo que la única diferencia entre los dos grupos sea la exposición en estudio?* La magnitud del riesgo relativo es una consideración importante (pero no la única –véase Capítulo 13) al establecer si una exposición en particular es una causa del resultado de interés.

Una vez hemos establecido que una exposición está asociada *causalmente* con el resultado de interés, es importante expresar la magnitud absoluta de su impacto sobre la aparición de la enfermedad en el grupo de expuestos (véase Apartado 5.2.2). Si tenemos información sobre el riesgo (o tasa) habitual de una determinada enfermedad en ausencia de la exposición, así como en su presencia, podremos determinar el *riesgo en exceso* (también conocido como *riesgo atribuible*) asociado con la exposición.

Riesgo en exceso = riesgo (o tasa) en los expuestos – riesgo (o tasa) en los no expuestos

Es útil expresar el riesgo en exceso con relación al riesgo total de la enfermedad entre los expuestos al factor en estudio. Esta medida se denomina *fracción de exceso* (también conocida como *porcentaje de riesgo en exceso* o *porcentaje de riesgo atribuible*) y describe la proporción de enfermedad en el grupo de expuestos que es atribuible a la exposición.

Fracción de exceso (%) = $100 \times (\text{riesgo en exceso} / \text{riesgo (o tasa) en los expuestos})$

Alternativamente, puede calcularse utilizando la fórmula siguiente:

Fracción de exceso (%) = $100 \times (\text{riesgo relativo} - 1) / \text{riesgo relativo}$

Ejemplo 16.1. Supongamos que se realizó un estudio de cohorte en una determinada localidad de Minas Gerais (Brasil) para evaluar la relación entre el hábito de fumar y el cáncer de pulmón en hombres. Supongamos, para simplificar, que los fumadores y los no fumadores eran similares respecto a otros factores de riesgo para cáncer de pulmón tales como la edad y exposiciones ocupacionales. Los resultados se muestran en la Tabla 16.1.

	Fumadores	No fumadores	Todos
Número de casos	120	10	130
Personas-años	54.545	50.000	104.545
Tasa por 100.000 p-as	220	20	124

Tabla 16.1.
Incidencia de cáncer de pulmón en fumadores y no fumadores: datos hipotéticos.

En este estudio, la razón de tasas es

$220 \text{ por } 100.000 \text{ p-as} / 20 \text{ por } 100.000 \text{ p-as} = 11$

y el riesgo en exceso asociado al fumar (asumiendo causalidad) es

$\therefore \text{Riesgo en exceso} = 220 \text{ por } 100.000 \text{ p-as} - 20 \text{ por } 100.000 \text{ p-as} = 200 \text{ por } 100.000 \text{ p-as}$

Para evaluar que proporción de 200 por 100.000 p-as corresponde a la tasa entre fumadores (220 por 100.000 p-as), podemos calcular la fracción de exceso:

Fracción de exceso (%) = $100 \times (200 \text{ por } 100.000 \text{ p-as} / 220 \text{ por } 100.000 \text{ p-as}) = 91\%$

Ésta es la proporción de casos de cáncer de pulmón en fumadores que puede atribuirse a fumar.

La *fracción de exceso* responde a la pregunta: *¿Qué proporción de casos nuevos de enfermedad en los expuestos puede atribuirse a la exposición?* Otra forma de utilizar este concepto es pensar en el mismo como la disminución en la incidencia de una enfermedad si los expuestos nunca hubieran estado expuestos. Así, en el Ejemplo 16.1, teóricamente podrían haberse prevenido un máximo de 91% de los casos de cáncer de pulmón en fumadores si nunca hubieran fumado.

Si la exposición es *protectora*, pueden calcularse medidas análogas. Normalmente se denominan *reducción del riesgo* (también conocida como *riesgo prevenido*) y *fracción prevenida* (también conocida como *porcentaje de riesgo prevenido*).

Reducción del riesgo = riesgo (o tasa) en los no expuestos – riesgo (o tasa) en los expuestos

Ejemplo 16.2. *Se realizó un gran ensayo aleatorizado para estudiar el valor de un programa antitabaco en la reducción de la aparición de cáncer de pulmón. Al final del ensayo, la incidencia de cáncer de pulmón fue 155 por 100.000 p-as en los que recibieron la intervención y 240 por 100.000 p-as en los controles. Por tanto, el beneficio máximo conseguido por la intervención fue*

$$\begin{aligned} \text{Reducción del riesgo} &= 240 \text{ por } 100.000 \text{ p-as} - 155 \text{ por } 100.000 \text{ p-as} \\ &= 85 \text{ por } 100.000 \text{ p-as} \end{aligned}$$

Así, se previnieron 85 casos nuevos de cáncer de pulmón por 100.000 p-as mediante el programa antitabaco.

$$\text{Fracción prevenida (\%)} = 100 \times (85 \text{ por } 100.000 \text{ p-as} / 240 \text{ por } 100.000 \text{ p-as}) = 35\%$$

Por tanto, con el programa antitabaco se previnieron un 35% de los casos esperados de cáncer de pulmón entre fumadores.

Fracción prevenida (%) = $100 \times (\text{reducción del riesgo} / \text{riesgo (o tasa) en los no expuestos})$

La fracción prevenida tiende a ser considerablemente menor que la fracción de exceso (Ejemplo 16.2). Esto es debido a que generalmente es imposible eliminar completamente la exposición e, incluso aunque fuese posible, es poco probable que la incidencia de la enfermedad en los que dejan de estar expuestos disminuya al nivel de los que nunca han sido expuestos.

El cálculo del riesgo en exceso (o reducción del riesgo) requiere información sobre la incidencia de la enfermedad en los grupos expuestos y no expuestos. Esta información está disponible directamente en estudios de cohorte y estudios de intervención. No obstante, para estudios caso-con-

trol, no puede calcularse el riesgo en exceso utilizando la fórmula indicada anteriormente porque no se conoce la incidencia de la enfermedad en los grupos expuestos y no expuestos. En estos casos, podemos calcular la fracción de exceso utilizando la fórmula basada en el riesgo relativo, que en estos estudios se estima por la *odds ratio*. En estudios caso-control poblacionales pueden utilizarse fórmulas alternativas para calcular el riesgo en exceso. Éstas se presentarán en el Anexo 16.1.

Medidas de impacto poblacional

Las medidas del efecto comentadas hasta ahora comparaban la incidencia en el grupo de *expuestos* con la incidencia de la enfermedad en el grupo de no expuestos. Para evaluar la incidencia adicional de la enfermedad en toda la *población en estudio* que puede atribuirse a la exposición, puede calcularse una medida denominada el *riesgo en exceso de la población* (también conocido como *riesgo atribuible en la población*). Ésta se define como

Riesgo en exceso de la población = riesgo (o tasa) en la población – riesgo (o tasa) en los no expuestos

o, asimismo, como

Riesgo en exceso de la población = riesgo en exceso \times proporción de la población expuesta al factor

Ejemplo 16.3. Volviendo al estudio hipotético descrito en el Ejemplo 16.1, la proporción de fumadores en la cohorte fue del 52%. Si este 52% de la población en estudio que fumaba no hubiera fumado nunca, su incidencia de cáncer de pulmón se habría reducido de 220 a 20 casos por 100.000 p-as.

Riesgo en exceso de la población = (220 por 100.000 p-as – 20 por 100.000 p-as) \times 0,52 = 104 por 100.000 p-as

Asimismo, el riesgo en exceso de la población en estudio puede calcularse restando la tasa en el grupo de no expuestos de la tasa en la población total en estudio. La tasa en la población total en estudio fue 124 por 100.000 p-as. (Tabla 16.1). Por tanto,

Riesgo en exceso de la población = 124 por 100.000 p-as – 20 por 100.000 p-as = 104 por 100.000 p-as

Por tanto, 104 casos de cáncer de pulmón por 100.000 p-as podrían haberse prevenido en toda la población en estudio si ninguno de los fumadores hubiera fumado.

Al igual que el riesgo en exceso entre los individuos expuestos, el *riesgo en exceso de la población* es una medida del riesgo de la enfermedad en la *población en estudio* que es atribuible a una exposición. Podemos expresar el riesgo en exceso de la población en relación con el riesgo total de la enfermedad en toda la población. Esta medida es la *fracción de exceso de la población* (también conocida como *fracción atribuible de la población*).

$$\begin{aligned} \text{Fracción de exceso de la población (\%)} &= \\ &= 100 \times \frac{\text{Riesgo en exceso de la población}}{\text{Tasa (o riesgo) en la población total}} \end{aligned}$$

Alternativamente, puede calcularse utilizando la fórmula siguiente

$$\begin{aligned} \text{Fracción de exceso de la población (\%)} &= \\ &= 100 \times \frac{P_e (\text{riesgo relativo} - 1)}{P_e (\text{riesgo relativo} - 1) + 1} \end{aligned}$$

donde p_e representa la prevalencia de la exposición en la población en estudio.

Ejemplo 16.4. En el Ejemplo 16.3, la fracción de exceso de la población sería

$$\begin{aligned} \text{Fracción de exceso de la población (\%)} &= 104 \text{ por } 100.000 \text{ p-as} / 124 \text{ por } 100.000 \text{ p-as} \\ &= 84\%. \end{aligned}$$

Esto significa que (asumiendo causalidad) aproximadamente un 84% de la incidencia de cáncer de pulmón en la población en estudio es atribuible al fumar. Por tanto, un 84% de los casos de cáncer de pulmón en esta población se habrían prevenido si los fumadores no hubieran fumado nunca.

La fracción de exceso de la población es una medida importante. Responde a la pregunta: *¿Qué proporción (fracción) de los casos nuevos de la enfermedad observados en la población en estudio es atribuible a la exposición a un factor de riesgo?* Por tanto indica qué proporción de la experiencia de la enfermedad en la población podría prevenirse si la exposición al factor de riesgo nunca se hubiera producido.

Observe que la fracción de exceso de los expuestos siempre será mayor que la fracción de exceso de la población, ya que la población en estudio ya incluye algunas personas no expuestas que, obviamente, no pueden beneficiarse de la eliminación de la exposición.

A veces es útil calcular la fracción de exceso de la población para una población mucho mayor que la población en estudio. Por ejemplo, los planificadores de salud pública están especialmente interesados en utilizar datos

de estudios epidemiológicos realizados en subgrupos de la población para estimar la proporción de casos en una región o en un país que puede ser atribuible a una exposición determinada (Ejemplo 16.5). En este caso, es necesario obtener datos sobre la prevalencia de la exposición en estas poblaciones a partir de otras fuentes.

Ejemplo 16.5. Volviendo al ejemplo anterior, una encuesta de salud realizada recientemente en los hogares de la región de Minas Gerais mostró que la prevalencia de fumar en hombres es del 35%. Por tanto, la proporción de casos de cáncer de pulmón que se produjo en hombres en toda la región que puede atribuirse a fumar puede calcularse como

$$\text{Fracción de exceso de la población (\%)} = 100 \times \frac{0,35 \times (11 - 1)}{0,35 \times (11 - 1) + 1} = 78\%$$

Así, en este ejemplo hipotético, un 78% de los cánceres de pulmón en toda la población masculina de Minas Gerais podría atribuirse al fumar. Obsérvese que esto es menor que el valor para la propia población de estudio. La explicación es que la prevalencia de fumar era menor en toda la población masculina de Minas Gerais (35%) que en la población de estudio (52%).

La Tabla 16.2 muestra cómo varía la fracción de exceso de la población respecto al nivel de prevalencia de la exposición en la población en estudio (p_e) y la magnitud del riesgo relativo. Está claro que la proporción de casos en una población concreta que puede atribuirse a una determinada exposición depende del valor del riesgo relativo y de la prevalencia de la exposición en la población. Por ejemplo, el fumar tabaco, con un riesgo relativo de alrededor de 5, y la exposición ocupacional a aminas aromáticas, con un riesgo relativo de alrededor de 500, están claramente implicados como causas de cáncer de vejiga. A pesar del hecho de que el riesgo relativo es mucho menor para fumar que para las aminas aromáticas, la fracción de exceso de la población es considerablemente mayor para fumar porque esta exposición es mucho más habitual que la exposición a aminas aromáticas. Se ha estimado que la fracción de exceso de la población para fumar en Inglaterra es del 46% en hombres (Morrison *et al.*, 1984) mientras que la fracción de exceso de la población para todas las exposiciones ocupacionales

Prevalencia de la exposición (p_e) (%)	Riesgo relativo		
	2	5	10
10	0,09	0,29	0,47
25	0,20	0,50	0,69
50	0,33	0,67	0,82
75	0,43	0,75	0,87
95	0,49	0,79	0,90

Tabla 16.2.

Fracciones en exceso de la población para diferentes niveles de prevalencia de la exposición y varios valores del riesgo relativo.

(incluyendo la exposición a aminas aromáticas) se encuentra sólo entre el 4 y el 19% (Vineis y Simonato, 1986). Por tanto, podría prevenirse un número mucho mayor de casos de cáncer de vejiga eliminando el fumar que eliminando las exposiciones ocupacionales.

Estas medidas de impacto poblacional presentan una serie de limitaciones. En primer lugar, se tiene que suponer que el factor de riesgo está asociado causalmente con la enfermedad de interés. En el Capítulo 13 se presenta una discusión sobre los criterios que pueden utilizarse para evaluar si es verosímil que una asociación observada sea causal. En segundo lugar, debe suponerse que no existe confusión o sesgo en las medidas de la incidencia entre los grupos de expuestos y no expuestos. Hasta ahora en nuestra discusión, para simplificar, hemos supuesto que los *grupos de expuestos y no expuestos eran similares excepto para la exposición en estudio*. Esto se produce raramente excepto en los grandes ensayos aleatorizados de intervención. En nuestros ejemplos anteriores, por ejemplo, deberíamos haber tenido en cuenta las diferencias en la distribución de la edad entre fumadores y no fumadores. Esto puede realizarse utilizando técnicas similares a las descritas en el Capítulo 14 para calcular *las medidas ajustadas*. A continuación, pueden utilizarse estas medidas ajustadas para calcular las medidas absolutas del efecto y las medidas del impacto poblacional utilizando las mismas fórmulas anteriores. En tercer lugar, hay que recordar que las estimaciones del riesgo relativo generalmente se obtienen a partir de estudios caso-control, de cohorte o de intervención. Estos estudios con frecuencia se realizan en subgrupos especiales de la población tales como inmigrantes, trabajadores manuales, etc. No obstante, los niveles de exposición y la susceptibilidad intrínseca en estos subgrupos pueden ser bastante diferentes de los de la población general. Por tanto es importante que la extrapolación de los datos de estos estudios a otras poblaciones se realice con cierta precaución. Por ejemplo, muchos estudios de cohorte se basan intencionadamente en grupos con unos niveles de exposición mucho más altos que los de la población general (p.e. cohortes ocupacionales) y los riesgos relativos obtenidos a partir de los mismos no deben utilizarse como tales para proporcionar estimaciones de la fracción de exceso de la población para otras poblaciones con niveles de exposición mucho menores. Esto puede evitarse si los niveles de exposición se miden con mayor precisión (en lugar de sólo 'expuestos' frente a 'no expuestos') y las estimaciones de las fracciones en exceso de la población tienen en cuenta las prevalencias de los distintos niveles de exposición.

Se puede calcular la proporción de un cáncer particular en una población determinada que está causado por la dieta, alcohol, tabaco, etc. Estos porcentajes pueden sumar hasta más del 100%. Esto es debido a que cada cálculo individual de la fracción de exceso de la población no tiene en cuenta el hecho de que estos factores de riesgo interaccionan entre sí. Por ejemplo, al calcular la proporción del cáncer de laringe debida a fumar, ignoramos el hecho de que algunos cánceres que se produjeron en fumadores sólo se produjeron porque también estaban expuestos al alcohol.

16.2.2 Papel y evaluación de medidas preventivas primarias

Una vez identificados los factores de riesgo y estimado su impacto en la población, es importante considerar los métodos para eliminar o reducir la exposición a los mismos. La prevención primaria incluye dos estrategias que con frecuencia son complementarias. Puede centrarse en toda la población, con el objetivo de reducir el riesgo promedio (la *estrategia poblacional*), o en las personas de alto riesgo (la *estrategia del individuo de alto riesgo*). Aunque la estrategia del individuo de alto riesgo, que está destinada a proteger individuos susceptibles, es la más eficaz para las personas con el mayor riesgo de un cáncer específico, estas personas pueden contribuir poco al impacto global de la enfermedad en la población. Por ejemplo, los pacientes con trasplante de órganos son especialmente susceptibles de desarrollar cáncer de piel no-melanoma (Bouwes Bavinck *et al.*, 1991). El tumor tiende a desarrollarse en las partes del cuerpo más expuestas al sol. Las campañas de prevención primaria para pacientes con trasplante de órganos tales como reducción de la exposición al sol y uso de protectores solares es probable que sean muy beneficiosas para estos pacientes, pero tendrán poco impacto sobre la magnitud global de la enfermedad en la población debido a que los pacientes con trasplante de órganos representan una proporción muy pequeña de la población. En este caso, debe aplicarse la estrategia de población o una combinación de las dos estrategias.

La principal ventaja de la estrategia poblacional es que puede producir los mayores beneficios a nivel poblacional y no requiere la identificación del grupo de alto riesgo. Su principal inconveniente es que requiere la participación de grandes grupos de personas para el beneficio de relativamente unos pocos. Por ejemplo, la adopción, por una población, de medidas para reducir la exposición solar, puede reducir el riesgo de cáncer de piel a nivel poblacional pero presentará un beneficio aparente pequeño para la mayoría de los individuos ya que la enfermedad es rara incluso entre los expuestos. Este fenómeno se denomina la *paradoja de prevención* (Rose, 1985).

Se han utilizado varias aproximaciones para reducir o eliminar la exposición a un factor de riesgo en particular, algunos ejemplos de los cuales se presentan en el Cuadro 16.1.

Si se han tomado medidas preventivas específicas para reducir la incidencia de un determinado cáncer, será necesario establecer si el esfuerzo ha tenido algún efecto positivo. La evaluación de los esfuerzos preventivos primarios a nivel poblacional se realiza principalmente en términos de monitorización de cambios en los riesgos de cáncer respecto a los cambios en la exposición a factores de riesgo. Por ejemplo, las tendencias temporales en la incidencia de un cáncer pueden compararse con cambios temporales en la exposición a un factor de riesgo en particular para demostrar si se está consiguiendo el efecto deseado. Esto se ilustra en la Figura 16.2, que muestra las tendencias en consumo de cigarrillos per cápita en los EE.UU. respecto al tiempo de implementación de iniciativas de control del tabaco, acontecimientos históricos importantes y tendencias en la mortalidad por cáncer de pulmón.

Cuadro 16.1. Ejemplos de estrategias utilizadas para reducir o eliminar la exposición a un factor de riesgo.

- Educación sanitaria individual o comunitaria (p.e. campañas en medios de comunicación social promocionando el uso de protectores solares).
- Regulación de carcinógenos en situaciones ocupacionales y en el medio ambiente (p.e. mejora de la protección contra la radiación).
- Regulación de precios (p.e. imponer impuestos en las compras de tabaco y alcohol).
- Anunciar restricciones (p.e. anuncios de prohibición de fumar o obligar la impresión de advertencias sanitarias en los paquetes de cigarrillos).
- Restricciones de tiempo y lugar del consumo (p.e. prohibir fumar en lugares públicos).

Los siguientes aspectos deben tenerse en cuenta al interpretar las tendencias de incidencia respecto a los cambios en la exposición a factores de riesgo. En primer lugar, si la tendencia empezó a disminuir antes de la introducción de la medida preventiva, es difícil atribuir una disminución reciente en la incidencia a la medida preventiva específica en estudio. En segundo lugar, debido al largo periodo de inducción del cáncer, pueden pasar muchos años o incluso décadas antes de que se observe algún efecto de una medida preventiva en las tendencias de incidencia o mortalidad (excepto si el factor de riesgo actúa en las últimas etapas de la carcinogénesis), como se muestra en la Figura 16.2. En tercer lugar, si las medidas tomadas están dirigidas sólo a una pequeña fracción de la población (p.e. una región), la evaluación debe limitarse a la misma subpoblación, de lo contrario, el efecto puede pasar desapercibido. En cuarto lugar, cuando la intervención ha sido limitada a una zona, pueden realizarse comparaciones de los cambios acontecidos en la zona de intervención frente a los acontecidos en las áreas «control».

Existen relativamente pocos estudios en sujetos que han adoptado hábitos o estilos de vida potencialmente más sanos y principalmente se han centrado a la investigación del riesgo de cáncer en exfumadores. Estos estudios muestran una clara disminución del riesgo, relacionada con el tiempo desde que se dejó de fumar y constituyen la evidencia más potente de la eficacia de dejar de fumar en la prevención del cáncer (Rose y Colwell, 1992; Gupta *et al.*, 1990). Más recientemente, el potencial de otros cambios en el estilo de vida (p.e. cambios en la dieta) en la prevención primaria del cáncer también se ha evaluado (o se está evaluando) en grandes ensayos de intervención (p.e. *Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group*, 1994; Hennekens *et al.*, 1996; Omenn *et al.*, 1996; Chlebowski y Grosvenor, 1994).

Los estudios basados en el individuo también han ayudado a evaluar la eficacia de las medidas preventivas en el lugar de trabajo. Por ejemplo, las mujeres empleadas por primera vez antes de 1925 en la industria de pintura con radio de agujas de reloj en los Estados Unidos presentaban unos riesgos muy aumentados de mortalidad por cáncer de huesos y de leucemia y otras enfermedades hematológicas, pero los riesgos disminuían en aquellas empleadas en los años subsiguientes (Tabla 16.3). La reducción del riesgo coincidía con cambios en las regulaciones de trabajo en la industria, que incluyeron la prohibición de tocar o afilar los pinceles con los labios en 1925-26 (Figura 16.3). Estas medidas redujeron considerablemente la exposición de las trabajadoras al radio.

Se están evaluando otras estrategias a la prevención primaria tales como el uso de campañas de vacunación masiva (p.e. vacuna antihepatitis B contra el cáncer de hígado (*The Gambia Hepatitis Study Group*, 1987)) y la quimiopreprofilaxis (p.e. tamoxifeno en la prevención de cáncer de mama en mujeres de alto riesgo (Powles *et al.*, 1994)).

16.3 Prevención secundaria

La prevención secundaria se refiere a la detección de cáncer en una etapa inicial, cuando el tratamiento es más eficaz que en el momento del diagnóstico y tratamiento habituales. Con estas medidas a veces puede prevenirse la progresión de la enfermedad y sus complicaciones (incluyendo la muerte).

16.3.1 Cribaje

El cribaje representa un componente importante de la prevención secundaria. Implica la aplicación de una prueba relativamente simple y barata a sujetos asintomáticos con objeto de clasificarlos como portadores probables o no de la enfermedad objeto del cribaje. Los casos positivos pueden someterse a subsiguientes procedimientos convenciona-

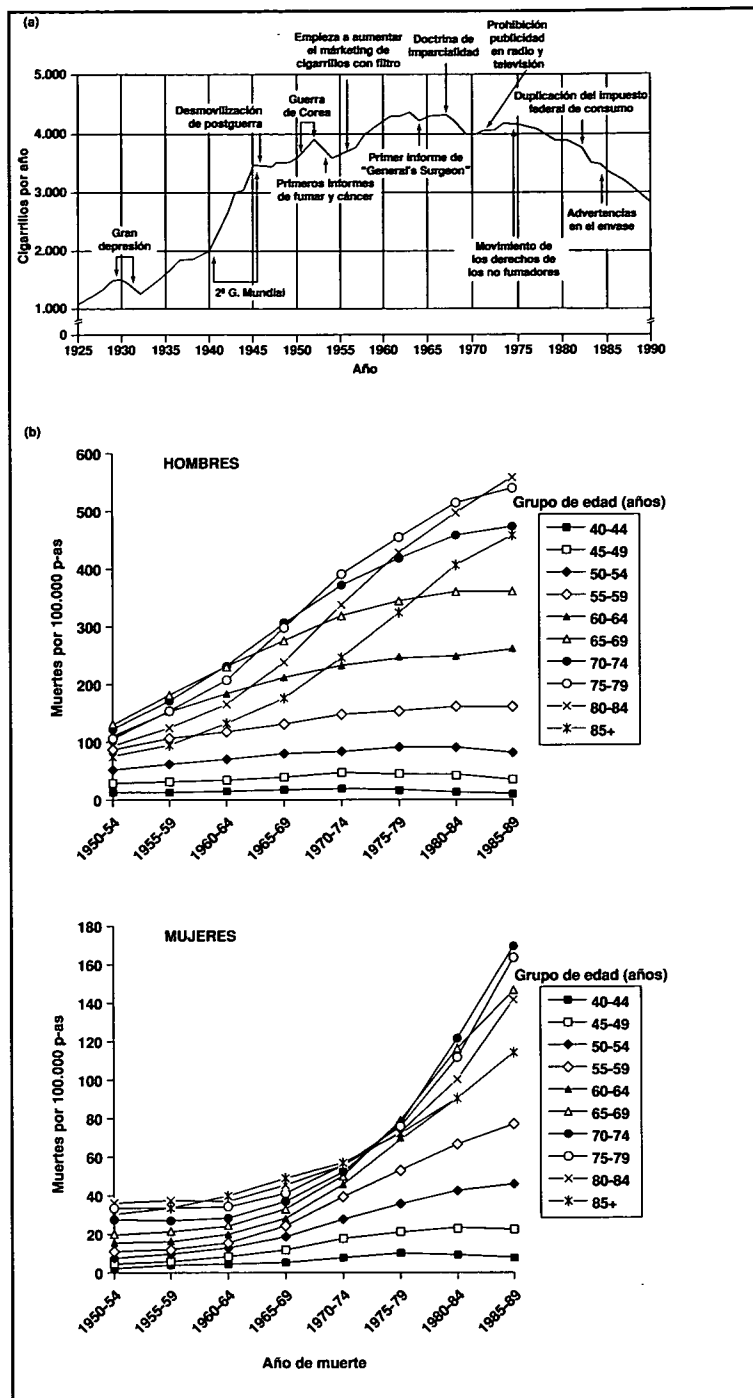


Figura 16.2.

(a) Consumo de cigarrillos per cápita entre personas de 18 años o más, Estados Unidos, 1925-1990 (reproducido con la autorización de *U.S. Department of Health and Human Services*, 1991) y (b) tendencias de mortalidad por cáncer de pulmón específicas de edad. Estados Unidos, 1950-1989 (reproducido con la autorización de Gilliland y Samet, 1994).

Tabla 16.3.

Mortalidad de mujeres empleadas en la industria de pintura con radio de agujas de reloj y seguida hasta el final de 1975, por año de primer empleo^a.

Año del primer empleo	Cáncer de huesos		Leucemia y enfermedades hematológicas	
	RME (O/E) ^b	Nº de casos observados (O)	RME (O/E) ^b	Nº de casos observados (O)
1915-19	233**	7	7,4*	2
1920-24	154**	20	3,3*	4
1925-29	10	1	1,0	1

^a Datos de Polednak *et al.* (1978)
^b Cifras esperadas obtenidas de las tasas de mortalidad por causas específicas para mujeres blancas de los Estados Unidos.
 RME=razón de mortalidad estandarizada.
 * $P < 0,05$; ** $P < 0,001$.



Figura 16.3. Ilustración de un periódico de Nueva York aludiendo a la intoxicación por radio de las pintoras de agujas de reloj (reproducido en Mould (1983)).

les de diagnóstico y, si es necesario, a un tratamiento adecuado. Las actividades de cribaje se basan en la asunción de que la detección y tratamiento precoces retrasarán o pararán la progresión de los casos de enfermedad establecidos, cuando el tratamiento iniciado más tarde puede ser menos eficaz. El principal objetivo del cribaje para un determinado cáncer es reducir la mortalidad por esta enfermedad entre los sujetos incluidos en el cribaje.

No obstante, el concepto de cribaje no es tan sencillo como puede parecer al principio. El tratamiento precoz no siempre mejora el pronóstico e, incluso si lo hace, los beneficios reales de cualquier tipo de cribaje deben evaluarse respecto a sus riesgos y costes y respecto a los beneficios que pueden derivarse de otras actividades de salud pública. El valor final de todo programa de cribaje sólo puede establecerse tras un proceso estricto de evaluación.

Toda actividad de cribaje de cáncer requiere (1) una enfermedad adecuada, (2) una prueba adecuada y (3) un programa de cribaje adecuado.

Enfermedad adecuada

Para ser adecuada para el control por un programa de detección y tratamiento precoces, una enfermedad tiene que incluir una fase preclínica durante la cual sea detectable (véase Figura 16.1), y el tratamiento precoz debe ofrecer alguna ventaja frente al tratamiento posterior. Obviamente, no tiene sentido realizar un cribaje de una enfermedad que no pueda detectarse antes de que la aparición de los síntomas lleven a la consulta médica y, si el tratamiento precoz no es especialmente útil, no tiene sentido una detección precoz.

Fase preclínica detectable

La fase preclínica de un cáncer empieza con el inicio biológico de la enfermedad (punto A en la Figura 16.1). La enfermedad progresa a continuación y alcanza un punto en el que puede detectarse por la prueba de cribaje (punto B en la Figura 16.1). A partir de este punto en adelante se dice que la fase preclínica de la enfermedad es «detectable». El punto de inicio de esta fase preclínica detectable depende, en parte, de las características del individuo y, en parte, de las características de la prueba utilizada. Una prue-

ba que puede detectar un estado muy 'precoz' del cáncer está asociada a una más larga fase preclínica detectable que una prueba que puede detectar sólo lesiones más avanzadas.

La proporción de una población que tiene una enfermedad preclínica detectable – su prevalencia – es un determinante importante de la utilidad del cribaje en el control de la enfermedad. Si la prevalencia es muy baja, se detectarán pocos casos, insuficientes para justificar los costes del programa de cribaje. En el momento del cribaje inicial, la prevalencia de la fase preclínica se determina por su incidencia y su duración media (recuérdese la discusión sobre la prevalencia en el Apartado 4.2). No obstante, en exámenes de cribaje subsiguientes, la prevalencia de la fase preclínica se determina principalmente por su incidencia, siendo la duración relativamente poco importante si el intervalo entre los exámenes es corto. Por tanto, la prevalencia de la fase preclínica y el número de casos detectados por el programa es máximo al final del primer examen de cribaje, mientras que cuanto más corto es el intervalo entre exámenes, menor es el número de casos detectado por examen (y mayores los costes por caso detectado).

Tratamiento precoz

Para que el cribaje sea beneficioso, el tratamiento administrado durante la fase preclínica detectable debe dar lugar a un mejor pronóstico que el tratamiento administrado después de que se desarrollen los síntomas. Por ejemplo, el cáncer de cuello uterino se desarrolla lentamente, requiriendo probablemente más de una década para las células cancerosas, que inicialmente están confinadas en la parte superficial del cuello uterino, para progresar a una fase de invasión. Durante esta etapa preinvasiva, el cáncer suele ser asintomático pero puede detectarse por cribaje utilizando el frotis Papanicolaou. El pronóstico de la enfermedad es mucho mejor si el tratamiento empieza durante esta etapa que cuando el cáncer ya se ha convertido en invasivo.

Por otro lado, si el tratamiento precoz no supone una diferencia porque el pronóstico es igualmente bueno (o igualmente malo) ya se inicie el tratamiento antes o después del desarrollo de los síntomas, la aplicación de una prueba de cribaje no será necesaria ni eficaz y puede incluso ser perjudicial (véase más adelante).

Importancia relativa de la enfermedad

Las tasas de prevalencia, incidencia o mortalidad por cáncer pueden utilizarse para evaluar si la enfermedad es un problema de salud pública lo suficientemente importante como para promover un programa de cribaje. Incluso en el caso de una enfermedad muy rara, si el pronóstico es muy grave y es fácilmente prevenible, todavía puede estar justificado realizar un cribaje. La consideración final dependerá inevitablemente de los beneficios, costes y la relación coste/beneficio respecto a las otras necesidades sanitarias de la población.

Prueba adecuada

Para que un programa de cribaje tenga éxito debe estar dirigido a una enfermedad adecuada con una prueba adecuada. Con objeto de definir una prueba de cribaje adecuada, es necesario considerar su validez y aceptabilidad.

Validez

La evaluación preliminar de una prueba de cribaje debe incluir estudios de su fiabilidad, la cual se evalúa como la variación intra- e inter-observador (véase Apartado 2.6). Aunque incluso una fiabilidad perfecta no asegura la validez, una prueba no fiable no será lo suficientemente válida para ser útil. Por otro lado, una prueba muy válida debe ser muy fiable.

Cuadro 16.2. Sensibilidad, especificidad y valor predictivo de una prueba de cribaje.

		Estándar de oro	
		Positivo	Negativo
Prueba de cribaje	Positiva	<i>a</i>	<i>b</i>
	Negativa	<i>c</i>	<i>d</i>

Verdaderos positivos = <i>a</i>	Falsos positivos = <i>b</i>
Verdaderos negativos = <i>d</i>	Falsos negativos = <i>c</i>

- La *sensibilidad* de la prueba de cribaje es la proporción de los individuos clasificados como positivos por el estándar de oro e identificados correctamente por la prueba de cribaje

$$\text{Sensibilidad} = a/(a+c)$$

- La *especificidad* de la prueba de cribaje es la proporción de aquellos clasificados como negativos por el estándar de oro e identificados correctamente por la prueba de cribaje:

$$\text{Especificidad} = d/(b+d)$$

- El *valor predictivo de un resultado positivo de una prueba de cribaje* representa la probabilidad de que alguien con un resultado positivo de la prueba de cribaje tenga realmente la enfermedad

$$\text{Valor predictivo de una prueba de cribaje positiva} = a/(a+b)$$

- El *valor predictivo de un resultado negativo de una prueba de cribaje* representa la probabilidad de que alguien con un resultado negativo de la prueba de cribaje no tenga realmente la característica de interés

$$\text{Valor predictivo de una prueba de cribaje negativa} = d/(c+d)$$

La validez de la prueba de cribaje puede expresarse por su sensibilidad y especificidad. Estas miden la capacidad de la prueba de identificar correctamente a las personas enfermas y a las no enfermas (véase en el Cuadro 16.2 y el Apartado 2.6 una revisión de estos conceptos). En la situación ideal, las pruebas presentarían una sensibilidad del 100% y una especificidad del 100%. En esta situación perfecta, la distribución de valores para la prueba de cribaje en los individuos con la enfermedad está completamente separada de la distribución de los valores en individuos sanos. En la Figura 16.4(a), todo afectado por la enfermedad tendrá un resultado de la prueba superior a un valor de corte elegido situado entre estas dos distribuciones. Lamentablemente, la realidad es más compleja. En general, las distribuciones de los resultados de la prueba de cribaje en los sujetos no enfermos y los enfermos se superponen (Figura 16.4(b)). En estas situaciones, la decisión del valor de corte para clasificar los resultados de la prueba de cribaje como positivos o negativos es arbitraria. En la Figura 16.4(b), algunos sujetos afectados de la enfermedad, con valores inferiores al valor de corte elegido se perderán en la prueba de cribaje. Por el contrario, algunas personas no enfermas se clasificarán erróneamente como positivas. Por tanto, existe en general un equilibrio entre la sensibilidad y especificidad de una prueba de cribaje determinada. La capacidad de detectar cuantos más casos posibles (sensibilidad alta) sólo puede aumentarse a expensas de un aumento del número de individuos sin la enfermedad que se clasificarán erróneamente como positivos por la prueba de cribaje (especificidad baja) y viceversa (ver un ejemplo numérico de esto en el Apartado 2.6.1).

La sensibilidad es un indicador del rendimiento en la identificación de casos (es decir, el número de casos realmente enfermos identificados por el programa) mientras que la especificidad es un indicador del número de falsos positivos. Aunque en un programa de cribaje sería deseable detectar a todos los sujetos con la enfermedad, utilizando una prueba con una sensibilidad máxima, esta opción puede dar lugar a una especificidad inaceptablemente baja. Los costes serían altos debido a la necesidad de seguir investigando los numerosos falsos positivos, con una sobrecarga de servicios de diagnóstico y el riesgo de desmotivar a los sujetos para participar en exámenes de cribaje subsiguientes. Por consiguiente, la elección del punto de corte depende de los costes relativos de los falsos positivos y los falsos negativos.

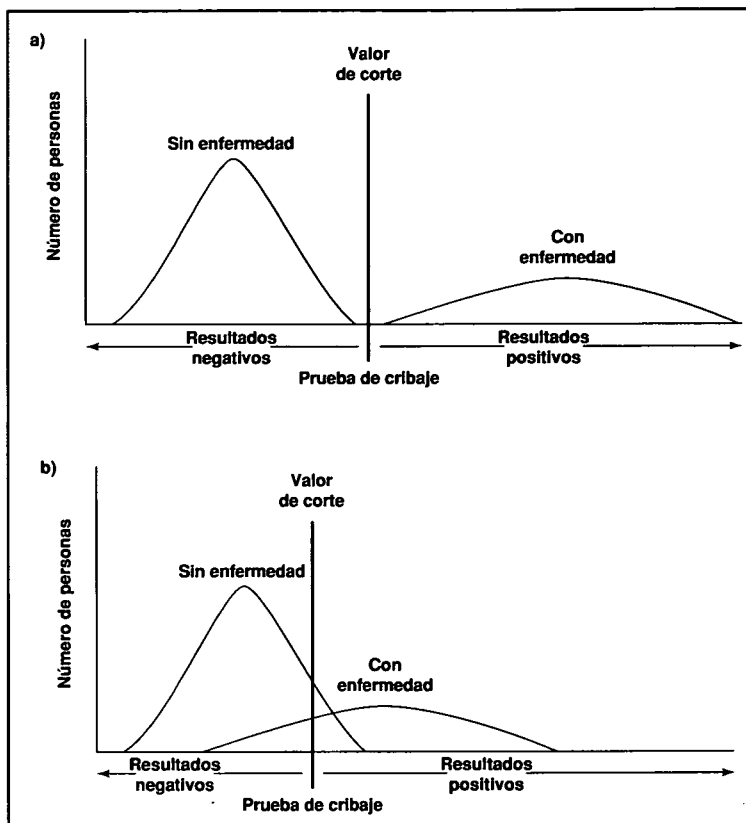


Figura 16.4.

Distribución de resultados de una prueba de cribaje en individuos no enfermos y enfermos: a) distribución ideal sin superposición; b) superposición de las distribuciones con el equilibrio inevitable entre sensibilidad y especificidad. En este ejemplo, en el que la prueba de cribaje tiende a proporcionar valores mayores a las personas con la enfermedad que a aquellas sin la enfermedad, el desplazamiento del valor de corte a la izquierda (e.d., disminuyendo su valor) aumenta la sensibilidad pero disminuye la especificidad; el desplazamiento a la derecha (e.d., aumentando su valor) aumenta la especificidad pero disminuye la sensibilidad.

No obstante, en la práctica es difícil estimar la sensibilidad de la prueba ya que no es posible aplicar una 'prueba estándar de oro' a la población cribada para determinar el número total de sujetos enfermos ($a+c$ en el Cuadro 16.2). La prueba de cribaje sólo nos proporciona el valor de a , es decir, el número de personas con una prueba de cribaje positiva y que se confirmó presentaban la enfermedad después de la evaluación del diagnóstico. La estrategia habitual para estimar la sensibilidad es seguir a los sujetos que son negativos (normalmente durante un año) para observar cuántos cánceres se desarrollarán entre ellos. Estos casos del 'intervalo' se consideran falsos negativos (c). La sensibilidad de la prueba de cribaje puede calcularse de la manera habitual. Sin embargo, el valor de esta estrategia es limitado ya que es difícil conseguir un seguimiento completo y, además, algunos de los cánceres de 'intervalo' pueden haber sido verdaderos negativos en el momento de la prueba de cribaje (e.d. tumores de crecimiento muy rápido).

Es más fácil estimar la especificidad si el cribaje está dirigido a una enfermedad poco frecuente como el cáncer. Prácticamente todos los cribados (N) están exentos de la enfermedad y por tanto N se puede utilizar para estimar el número total de personas no afectadas por la enfermedad ($b + d$ en el Cuadro 16.2). Debido a que todos los sujetos con cribaje positivo se siguen investigando, también se conoce el número de falsos positivos (b en el Cuadro 16.2) y, por tanto, el número de verdaderos positivos (d en el Cuadro 16.2) se puede calcular como $N - b$. La especificidad se puede estimar entonces como $(N - b) / N$.

Aceptabilidad y costes

Además de tener una validez adecuada, una prueba de cribaje debe ser barata, aceptable, simple, lo menos dolorosa posible y no debe causar complicaciones. Muchas pruebas de cribaje cumplen estos criterios – el frotis Papanicolaou para cáncer de cuello uterino es un buen ejemplo. Por el contrario, incluso aunque el cribaje sigmoidoscópico puede dar lugar a una reducción de la mortalidad por cáncer de colon, es cuestionable que una prueba así sea aceptable debido a los gastos, la incomodidad y el riesgo de perforación del intestino.

Programa de cribaje adecuado

La aplicación organizada de las actividades de diagnóstico y tratamiento precoces en grupos grandes, con frecuencia se denomina *cribaje de masas* o *cribaje poblacional*, y la serie de procedimientos implicados se describe como *programa de cribaje*.

Un programa de cribaje incluye un componente diagnóstico y uno terapéutico. Los dos se incluyen en el término "programa de cribaje" porque la detección precoz debe continuarse con el tratamiento para que sea útil en el control de la enfermedad. El componente diagnóstico incluye el protocolo de cribaje y los procedimientos para la evaluación diagnóstica de las personas con pruebas positivas. El protocolo de cribaje debe especificar de manera precisa quién se tiene que incluir en el cribaje, a qué edad, con qué

frecuencia, con qué prueba, etc.; debe ser dinámico en lugar de fijo y debe intentar optimizar los beneficios potenciales del programa de cribaje. El componente terapéutico es el proceso por el cual los casos confirmados son tratados. También debe ser dinámico y estar regulado por procedimientos universales estrictos que ofrezcan el mejor tratamiento disponible a todos los casos identificados.

Un programa de cribaje es una empresa compleja que implica la aplicación de una prueba determinada a una población particular en una situación particular. Las circunstancias varían en diferentes países y no debe suponerse que un formato adecuado para un país sea aplicable a otro sin probarlo y evaluarlo antes con un diseño lo suficientemente riguroso. En el Cuadro 16.3 se presenta una lista de aspectos esenciales de un programa de cribaje organizado.

Cuadro 16.3. Aspectos esenciales de un programa de cribaje organizado.

- Existe una definición clara de la población diana.
- Los individuos a cribar son identificables (p.e., lista con nombres y direcciones de todos los individuos elegibles en la población diana).
- Se dispone de las medidas para asegurar una alta cobertura y cumplimiento (p.e. carta personal de invitación).
- Se dispone de instalaciones de campo adecuadas para la obtención del material de cribaje e instalaciones de laboratorio adecuadas para examinarlo.
- Existe un programa de control de calidad organizado para evaluar el material de cribaje y su interpretación.
- Existen instalaciones adecuadas para el diagnóstico y tratamiento apropiado de las lesiones neoplásicas confirmadas y para el seguimiento de los individuos tratados.
- Hay circuitos de pacientes cuidadosamente diseñados para el seguimiento de cualquier anomalía observada y para proporcionar información sobre las pruebas de cribaje con resultados normales.
- La evaluación y monitorización del programa global están organizadas de tal modo que puedan calcularse las tasas de incidencia y mortalidad por separado para los participantes y los no participantes en el programa, a nivel de la población diana total. Se debe establecer un control de calidad de estos datos epidemiológicos.

(Modificado de Hakama *et al.*, 1986)

16.3.2 Evaluación de programas de cribaje

Incluso después de haber confirmado que una enfermedad es apropiada para el cribaje y de disponer de una prueba válida, sigue estando poco claro si debe activarse un programa generalizado de cribaje para esta enfermedad en una población determinada. El objetivo principal de la evaluación de un programa de cribaje es ayudar a decidir si es recomendable como actuación de salud pública para controlar un determinado cáncer.

La evaluación de un programa de cribaje potencial implica la consideración de dos aspectos. En primer lugar, si la organización del programa propuesto es *factible y coste-efectiva* (bajo coste por caso detectado), y en segundo lugar, si es *efectiva* para reducir la magnitud de la enfermedad. Los dos deben considerarse detenidamente. La implementación de un programa de cribaje, incluso siendo coste-efectivo, no estará justificada si no cumple su objetivo de reducir la morbilidad y mortalidad en la población diana.

Medidas del proceso

La viabilidad, aceptabilidad y costes de un programa pueden evaluarse mediante *medidas del proceso*, las cuales están relacionadas con los aspectos administrativos y organizativos del programa. Como ejemplos podemos citar la identificación de la población diana, el número de personas examinadas, la proporción de la población diana examinada, las instalaciones para el diagnóstico/tratamiento en los servicios sanitarios, el funcionamiento del sistema de referencia y su cumplimiento, el coste total, el coste por caso detectado, etc. La principal ventaja de estas medidas del proceso es que se obtienen directamente y son útiles para monitorizar la actividad del programa. Su principal limitación es que no proporcionan indicaciones de si las personas cribadas presentan una menor mortalidad por el cáncer estudiado por el programa que aquellos que no se cribaron.

Una medida del proceso especialmente útil es el valor predictivo de una prueba positiva (VPP). El valor predictivo positivo representa la proporción de personas que se confirman tienen la enfermedad en cuestión tras seguir la evaluación del diagnóstico de todos los individuos que fueron positivos para la prueba de cribaje ($a+b$ en el Cuadro 16.2). Un VPP alto indica que una proporción considerablemente alta de los costes de un programa se está gastando para la detección de la enfermedad durante su fase preclínica. Un VPP bajo indica que una gran proporción de los costes se está desperdiciando en la detección y evaluación del diagnóstico de falsos positivos, personas que tienen una prueba de cribaje positiva pero no presentaron la enfermedad en la investigación diagnóstica subsiguiente. No obstante, es importante destacar que el VPP es una medida proporcional; puede obtenerse un VPP alto incluso cuando la frecuencia de detección de casos es inaceptablemente baja. Por ejemplo, el VPP puede ser del 80% indicando que el 80% de los que fueron positivos en el cribaje eran realmente enfermos. No obstante, si sólo 10 sujetos son positivos en el cribaje, el número de casos detectados por el programa será sólo de 8. La principal ventaja de esta medida es que se obtiene poco tiempo después de iniciarse el pro-

grama de cribaje y, a diferencia de la sensibilidad, no es necesario un seguimiento para estimarla.

El valor predictivo positivo de una prueba de cribaje depende del número de verdaderos positivos a y del número de falsos positivos b (véase el Cuadro 16.2). Por tanto, el VPP puede aumentarse bien aumentando el número de verdaderos positivos o bien disminuyendo el número de falsos positivos. El número de verdaderos positivos puede incrementarse *optimizando la prevalencia de la enfermedad preclínica detectable*, por ejemplo, espaciando los cribajes para mantener la prevalencia de la enfermedad preclínica en la población diana a un nivel más alto. El número de falsos positivos puede reducirse *aumentando la especificidad de la prueba*, es decir, cambiando el criterio de positividad o repitiendo la prueba de cribaje después de una prueba positiva. Un VPP bajo es más probable que sea a consecuencia de una especificidad baja que de una sensibilidad baja. La especificidad de una prueba es la que determina el número de falsos positivos que se detectan entre las personas sin la enfermedad, que son la gran mayoría de las estudiadas en prácticamente cualquier programa. La sensibilidad es menos importante para una enfermedad rara porque afecta a menos personas. Por el contrario, una pequeña pérdida de especificidad puede dar lugar a un gran aumento en el número de falsos positivos y a una gran disminución en el VPP.

Efectividad en reducir la mortalidad

El segundo, y definitivo, aspecto en la evaluación de un programa de cribaje es su efectividad en cuanto a la reducción de la morbilidad y mortalidad por la enfermedad estudiada. Incluso aunque un programa de cribaje identifique de forma exacta y económica un gran número de individuos con la enfermedad preclínica, tendrá poco valor de salud pública si el diagnóstico y tratamiento precoces no tienen un impacto sobre el desenlace último de estos casos.

La obtención de una estimación exacta de una reducción en la mortalidad requiere un seguimiento a largo plazo de grandes poblaciones. Por consiguiente, se han utilizado *medidas de resultado intermedias* tales como el estadio en el momento del diagnóstico y la supervivencia (tasa de letalidad) que pueden estar disponibles en los primeros años después del inicio de un programa de cribaje. Por ejemplo, en un programa de cribaje eficaz, la distribución de estadios de los cánceres detectados debe desplazarse hacia las fases menos avanzadas de la enfermedad y el riesgo de morir por cáncer (tasa de letalidad) debe ser menor para los casos detectados por cribaje que para los casos diagnosticados por síntomas.

No obstante, existen limitaciones críticas asociadas con el uso de estas variables intermedias. La ausencia de un cambio en los parámetros puede significar que el cribaje no está teniendo éxito pero no proporcionan una medida del proceso adecuada ya que están afectadas por una serie de sesgos tales como sesgo de duración, sesgo de adelanto diagnóstico y sesgo de sobrediagnóstico.

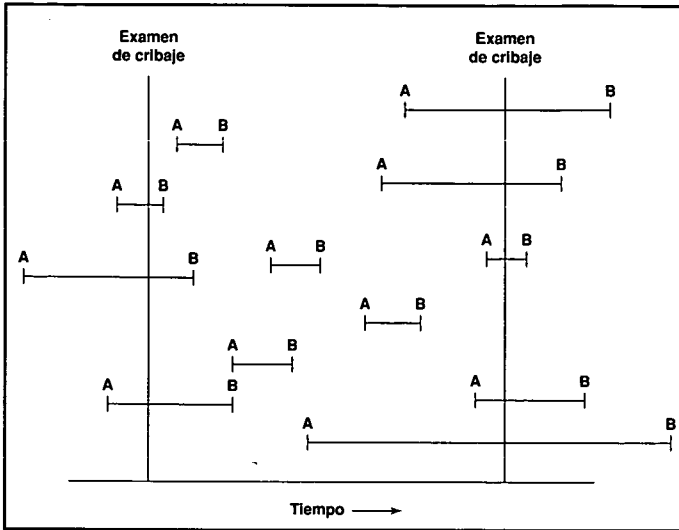


Figura 16.5. Diagrama que ilustra el sesgo de duración. Cada línea representa un caso desde el punto en que se hace detectable por la prueba de cribaje (A) hasta el punto en que aparecen síntomas/signos clínicos (B).

(a) *Sesgo de duración* – El sesgo de duración se refiere al fenómeno que se produce cuando los casos detectados por un programa de cribaje no son una muestra aleatoria de la distribución general de todos los casos de enfermedad preclínica en la población cribada. Esto es probable que se produzca cuando las pruebas de cribaje se aplican a intervalos moderadamente prolongados (es decir, una vez cada 2-5 años) de modo que sea más probable que se detecten los casos con una fase preclínica larga en lugar de aquellos con tumores de crecimiento más rápido (Figura 16.5). Por consiguiente, los casos detectados por cribaje pueden ser una muestra sesgada de todos los casos, sobrerrepresentando los casos con un pronóstico más

favorable, mientras que los casos con una fecha de inicio similar pero con evolución menos favorable se detectan por los síntomas clínicos. Por tanto, el sesgo de duración podría dar lugar a una conclusión errónea de que el cribaje era beneficioso cuando, de hecho, las diferencias observadas en la supervivencia (tasa de letalidad) eran meramente el resultado de la detección por cribaje de los casos menos rápidamente fatales, mientras que aquellos más rápidamente fatales se diagnosticaban en estadios más avanzados después del desarrollo de síntomas.

(b) *Sesgo de adelanto diagnóstico*– Si un individuo participa en un programa de cribaje y se le detecta la enfermedad antes de lo que se habría detectado en ausencia del cribaje, el tiempo en que se avanza el diagnóstico como resultado del cribaje es el tiempo de *adelanto diagnóstico*. Debido a que el cribaje se aplica a individuos asintomáticos, por definición, cada caso detectado por cribaje tendrá su diagnóstico adelantado un cierto tiempo. Si el adelanto diagnóstico es cuestión de días, meses o años, dependerá de la enfermedad, del individuo y del procedimiento de cribaje. Los casos que progresan rápidamente desde la enfermedad preclínica a la clínica ganarán menos tiempo de adelanto diagnóstico que los casos que se desarrollan más lentamente, con una fase preclínica más larga. El tiempo de adelanto diagnóstico también dependerá del intervalo entre el momento en que se realiza el cribaje y el momento en que la fase preclínica es detectable. Debido al fenómeno del adelanto diagnóstico, la fecha de diagnóstico se adelanta en el tiempo y la supervivencia medida desde el diagnóstico se prolonga automáticamente para los casos detectados por cribaje incluso cuando la duración total de la vida no aumente. Este fenómeno se conoce como *sesgo de adelanto diagnóstico*.

Supongamos que se cribaron 100 individuos para un cáncer X, un cáncer para el cual no existe un tratamiento eficaz, y que, en promedio, el cribaje identifica el cáncer 1 año antes de que sea clínicamente evidente. También se reunió un grupo similar de 100 personas a las que no se les ofrece nin-

gún tipo de cribaje. Estos dos grupos se siguieron durante cinco años y se detectaron cinco personas con el cáncer X en cada grupo.

Examinemos la supervivencia de los 5 casos en cada grupo. El curso de su enfermedad se muestra en la Figura 16.6. La supervivencia a 1,5 años para el grupo cribado es del 100% (todos estaban vivos 1,5 años después de ser cribados), mientras que la supervivencia a 1,5 años para el grupo no cribado fue sólo del 80% (1 caso murió 1 año después de la aparición de los síntomas), aunque los dos grupos tenían la misma duración total de vida (dada nuestra asunción inicial de ausencia de tratamiento eficaz).

El problema con este análisis es que el punto de inicio para monitorizar la supervivencia es diferente entre los casos cribados y los no cribados, siempre con el detrimento aparente de los casos detectados sin cribaje. El análisis adecuado es comparar la mortalidad de las 100 personas cribadas con la mortalidad de las 100 personas no cribadas *desde el momento del cribaje*. En el ejemplo anterior, la tasa de mortalidad en el grupo cribado es de 2 muertes en 496 personas-años (98 personas \times 5 años, más 1 persona \times 2 años, más 1 persona \times 4 años). La tasa es la misma en el grupo no cribado ya que el número de personas-años, contado desde el momento en que habría tenido lugar el cribaje si se hubiera realizado, es idéntico al del grupo cribado.

Existen dos formas de considerar el efecto del tiempo de adelanto diagnóstico sobre la evaluación de la eficacia de un programa de cribaje. La primera es comparar no la duración de la supervivencia desde el diagnóstico a la muerte, sino las tasas de mortalidad en los grupos cribado y no cribado (como se ha realizado en el ejemplo anterior). Como alternativa, si puede estimarse el tiempo de adelanto diagnóstico para una enfermedad, puede tenerse en cuenta y así poder comparar la supervivencia de los casos detectados por cribaje y por síntomas. Por ejemplo, el tiempo de adelanto diagnóstico medio para cáncer de mama se ha estimado en aproximadamente 1 año (Shapiro *et al.*, 1974). Por consiguiente, para evaluar la eficacia de un programa de cribaje de cáncer de mama, los riesgos de supervivencia a 2, 3, 4, 5 y 6 años de los casos cribados debe compararse, respectivamente, con los riesgos de supervivencia a 1, 2, 3, 4 y 5 años de los casos no cribados. No obstante, las determinaciones del tiempo de adelanto diagnóstico son difíciles de efectuar y no pueden generalizarse ya que dependen de la capacidad de cada programa de cribaje para detectar enfermedad preclínica.

(c) *Sobrediagnóstico* – Es posible que muchas de las lesiones detectadas por el programa de cribaje nunca hubiesen evolucionado a cáncer invasivo y muerte. Estas lesiones se conocen como pseudocáncer. Por tanto, el be-

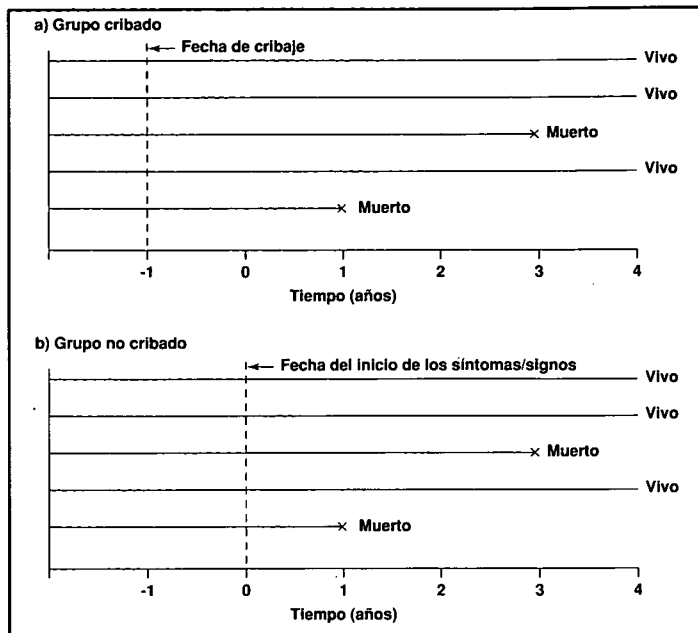


Figura 16.6. Diagrama que ilustra el sesgo de adelanto diagnóstico. Cada línea representa la experiencia de supervivencia para cada caso de cáncer en los grupos cribado y no cribado (ver texto).

neficio real de identificar lesiones preclínicas por cribaje puede ser mucho menor que el que se percibe.

En resumen, las medidas intermedias del resultado, tales como la distribución por estadios y la tasa de letalidad (supervivencia), pueden ser adecuadas como variables sucedáneas en un programa de cribaje, pero están sujetas al sesgo de adelanto diagnóstico, al sesgo de duración de la enfermedad y al sesgo de sobrediagnóstico. Por tanto, el principal parámetro para evaluar programas de cribaje destinados a detectar el cáncer en estadio inicial (p.e. cribaje de cáncer de mama y colon) es la reducción en la mortalidad. Cuando el cribaje está destinado a las condiciones precancerosas y a cánceres en estadios iniciales (p.e. cribaje de cáncer de cuello uterino), la reducción en la incidencia del cáncer invasivo y la reducción en la mortalidad son medidas adecuadas del resultado. Esto implica que todo programa de cribaje debe planificarse de tal forma que sea posible su evaluación en términos de reducción de la mortalidad (e incidencia) en la población total. A continuación se presenta una ilustración de cómo los resultados intermedios pueden inducir a error.

***Ejemplo 16.6.** Un total de 6.364 hombres fumadores de 40-64 años se distribuyeron aleatoriamente en un grupo de intervención que recibió un cribaje cada 6 meses por radiografía de tórax y citología de esputo durante 3 años y un grupo control que recibió un único examen al final del tercer año. Los casos de cáncer de pulmón detectados por cribaje se identificaron con un estadio menos avanzado, fueron resecables más frecuentemente y presentaban una supervivencia significativamente mejor que los casos detectados por síntomas. No obstante, no había diferencias significativas en la mortalidad entre los grupos de intervención y control (Kubik et al., 1990).*

En el ejemplo anterior, los resultados intermedios parecían indicar que el uso de radiografía de tórax y citología eran efectivos en el cribaje de cáncer de pulmón. No obstante, no se observó una reducción en la mortalidad. Resultados similares se han observado en todos los ensayos aleatorizados que han tratado este aspecto.

Estudios para evaluar la eficacia de un programa de cribaje Estudios de intervención

El ensayo aleatorizado es el mejor diseño de estudio para la evaluación de la eficacia de un programa de cribaje porque permite una evaluación experimental rigurosa. Cuando el tamaño de la muestra es lo suficientemente grande, el control de los factores de confusión está virtualmente garantizado por la aleatorización del proceso. La autoselección o el sesgo del voluntario, que es problemático para la comparación de los grupos cribado y no cribado en estudios observacionales, no puede influenciar la validez

de los resultados de los ensayos aleatorizados, ya que el programa de cribaje está asignado al azar por el investigador después de que los individuos hayan aceptado participar en el ensayo.

Ejemplo 16.7. *En Suecia, en 1977, se inició un ensayo aleatorizado controlado para evaluar la eficacia del cribaje masivo con mamografía única en la reducción de la mortalidad por cáncer de mama. Un total de 162.981 mujeres de 40 años o mayores residentes en las regiones de Kopparberg y Ostergotland se distribuyeron aleatoriamente para que se les ofreciera o no el cribaje cada 2 ó 3 años dependiendo de la edad. Los resultados hasta el final de 1984 mostraron que en las mujeres de 40 – 74 años, en el momento de la inclusión se observaba una reducción del 31% en la mortalidad por cáncer de mama en el grupo al que se ofreció el cribaje (Tabar et al., 1985).*

No obstante, existen varios problemas con los ensayos aleatorizados. En primer lugar, la contaminación del grupo control (el conocimiento del programa de cribaje puede llevar a los sujetos del grupo control a buscar el cribaje). En segundo lugar, puede ser necesaria una gran cantidad de sujetos en los ensayos de cribaje de enfermedades con una tasa de incidencia baja, como la mayoría de los cánceres, y/o cuando el ensayo está diseñado para mostrar pequeños beneficios (como en el Ejemplo 16.7). En tercer lugar, puede no ser aceptable la asignación aleatoria de sujetos a programas que no incluyan cribaje cuando ya se ha introducido un programa de cribaje en la población, incluso en ausencia de evidencia experimental (p.e. cribaje de cáncer de cuello uterino).

Estudios observacionales

Mientras que los ensayos aleatorizados pueden aportar la evidencia mejor y más válida en relación a la eficacia de un programa de cribaje, al igual que con la evaluación de hipótesis etiológicas, la mayor parte de la información procede de estudios no experimentales debido a los problemas inherentes relacionados con los costes, ética y viabilidad. Los estudios observacionales pueden proporcionar una información útil y necesaria, especialmente en las numerosas situaciones en que los ensayos aleatorizados no son posibles, tal como ocurre en el caso de procedimientos bien establecidos como el frotis de Papanicolaou. No obstante, la interpretación de los resultados de estos estudios no es tan sencilla.

Los *estudios ecológicos* se han utilizado para examinar tendencias en las tasas de enfermedad respecto a las frecuencias de cribaje dentro de una población, o para comparar la relación entre las frecuencias de cribaje y las tasas de enfermedad para diferentes poblaciones. Estos estudios son muy útiles para indicar la existencia de una relación entre cribaje y una disminución de la morbilidad o mortalidad. No obstante, deben tenerse en cuenta las limitaciones inherentes de los estudios ecológicos. En primer lugar, debido a que la información de estos estudios se refiere a poblaciones y no

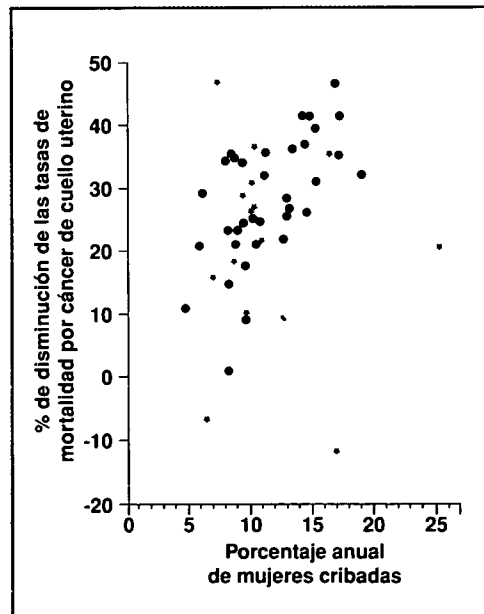
a individuos, no puede afirmarse que los que experimentan la reducción de la mortalidad sean las mismas personas que recibieron el cribaje (*falacia ecológica*; véase Apartado 11.2.2). Además, estos estudios no permiten el control de factores potenciales de confusión. Por último, la medida de la frecuencia de cribaje empleada normalmente es un valor medio para la población, de modo que no puede determinarse una estrategia de cribaje óptima para un individuo. Por tanto, los estudios ecológicos pueden sugerir un beneficio derivado de un programa de cribaje, pero no pueden demostrar esta hipótesis.

Los *estudios de cohorte* requieren el seguimiento a largo plazo de los sujetos cribados y no cribados. Al interpretar los resultados de estos estudios, deben tenerse en cuenta los efectos potenciales de la autoselección de los participantes ya que son especialmente importantes respecto a los factores de confusión así como los sesgos de adelanto diagnóstico y de duración.

En los estudios de cohorte, las personas sometidas a cribaje no se eligen aleatoriamente y los individuos que eligen ser cribados pueden diferir de los que se niegan al cribaje y de la población en general (sesgo de selección). Estos voluntarios pueden tener pronósticos totalmente diferentes en comparación con sus equivalentes no cribados. En general, los voluntarios tienden a tener una mejor salud y tasas de mortalidad menores que la po-

Figura 16.7. Cambio en la mortalidad por cáncer de cuello uterino de 1950-54 a 1965-69 respecto al porcentaje anual medio estimado de mujeres mayores de 19 años cribadas durante 1953-68, en cada estado de los EE.UU. Las estrellas indican los estados menos poblados (e.d. población femenina mayor de 19 años acumulada inferior a 5 millones durante el periodo de estudio) (reproducido con la autorización de Wiley-Liss, Inc., subsidiaria de John Wiley & Sons, Inc., de Cramer *et al.*, 1974).

Ejemplo 16.8. Para evaluar si las disminuciones recientes en la mortalidad por cáncer de cuello uterino en los EE.UU. estaban relacionadas con el cribaje de cuello uterino, se examinó el cambio en la mortalidad por cáncer de cuello uterino de 1950-54 a 1965-69 en los diferentes estados respecto a la proporción de mujeres cribadas en cada uno. Los resultados se presentan en la Figura 16.7. Había una correlación positiva entre la magnitud de la disminución de las tasas de mortalidad y el esfuerzo de cribaje en cada estado. Esta relación positiva se hace más evidente si se excluyen del análisis los estados menos poblados (coeficiente de correlación (r) = 0,60; valor $P < 0,0005$) (Cramer, 1974). Estos resultados eran compatibles con un efecto beneficioso de la citología de cuello uterino. No obstante, esta relación puede estar confundida por otros factores como los cambios socio-económicos.



blación general y es más probable que cumplan los regímenes médicos prescritos. Por otro lado, los voluntarios en un programa de cribaje pueden representar los preocupados sanos, es decir, individuos asintomáticos con un mayor riesgo de desarrollar la enfermedad debido a su historia médica o familiar, o a cualquier otra característica del estilo de vida. Estos individuos pueden presentar un riesgo de mortalidad aumentado independientemente de la eficacia del programa de cribaje. La dirección del sesgo de selección potencial puede ser difícil de predecir y todavía resulta más difícil cuantificar la magnitud de estos efectos.

***Ejemplo 16.9.** Para examinar el efecto protector del cribaje mediante frotis de Papanicolaou, se realizó un estudio de cohorte en dos regiones de Suecia (Uppsala y Gävleborg) donde se introdujo el cribaje citológico organizado en 1967 y 1972, respectivamente. Un total de 386.990 mujeres residentes en algún momento entre 1968 y 1992 en cualquiera de estas dos regiones se identificaron por el Registro de Población y se incluyeron en el estudio. La historia de cribaje de las mujeres se determinó por los registros informatizados existentes de frotis de Papanicolaou tomados en la zona. La interrelación de bases de datos permitió el seguimiento completo respecto a la incidencia de cáncer, la emigración fuera de las áreas de estudio y la supervivencia hasta 1992. Un total de 938 casos nuevos de cáncer de cuello uterino de células escamosas se produjeron durante el seguimiento de esta cohorte. Las mujeres que se habían cribado alguna vez presentaron aproximadamente la mitad del riesgo de aquellas que no se habían cribado nunca (razón de tasas = 0,55; intervalo de confianza del 95% = 0,51-0,61) (Sparén, 1996).*

Los estudios caso-control de cribaje implican la comparación de las historias de cribaje de sujetos que muestran, o no, el resultado que el cribaje pretende prevenir (muerte por cáncer o incidencia de enfermedad invasiva). Aunque los estudios caso-control se utilizan cada vez más para evaluar programas de cribaje (p.e. cáncer de cuello uterino), no pueden sustituir a los estudios experimentales porque son susceptibles de confusión y sesgo (por ejemplo, los casos pueden diferir de los controles en su capacidad de recordar la historia de cribaje en el pasado). No obstante, una vez se ha generalizado una forma de cribaje, los estudios caso-control pueden utilizar archivos existentes de modo que no aparecerá el sesgo de memoria. El primer estudio de este tipo comparó la historia de frotis Papanicolaou en 212 casos hospitalizados de cáncer de cuello uterino invasivo con controles vecinos agrupados por edad (Clarke y Anderson, 1979). Menos casos que controles se habían realizado un frotis Papanicolaou durante los 5 años anteriores al año de diagnóstico. El riesgo de cáncer de cuello uterino invasivo en mujeres que no se habían realizado un frotis Papanicolaou era aproximadamente tres veces el de las mujeres que se habían cribado después de ajustar por la situación socioeconómica. Los autores intentaron examinar el impacto del potencial sesgo de memoria comparando los datos obtenidos mediante

una muestra de entrevistas personales con los datos registrados en las historias clínicas. No hubo evidencias de que la información obtenida durante las entrevistas estuviera afectada por el sesgo de memoria.

Ejemplo 16.10. *En Cali (Colombia) se ofreció el cribaje de cáncer de cuello uterino a todas las mujeres sexualmente activas y se efectúa rutinariamente en clínicas prenatales desde finales de los años 60. Para evaluar el papel del frotis de Papanicolaou en la prevención del carcinoma de cuello uterino invasivo en esta población, se efectuó un estudio caso-control. Mediante el registro poblacional de cáncer se identificaron y entrevistaron con éxito un total de 204 casos nuevos de cáncer de cuello uterino invasivo durante los años 1977-1981. Para cada caso, se seleccionó un control del vecindario agrupado según el año de nacimiento del caso \pm 2 años. Los casos y controles se entrevistaron para obtener una historia de los frotis Papanicolaou efectuados con fines de cribaje. La historia sobre citología de cuello uterino se limitó al periodo de 12 a 72 meses antes de la fecha del diagnóstico del caso. Los exámenes realizados dentro de los 12 meses precedentes al diagnóstico se ignoraron porque era probable que estuvieran relacionadas con síntomas o con la enfermedad. Para cada control, el estudio cubrió el mismo intervalo de tiempo que el del caso correspondiente, determinado a partir de la fecha de diagnóstico. El riesgo de desarrollar carcinoma invasivo fue 10 veces mayor en las mujeres no cribadas que en las cribadas (Aristizabal et al., 1984).*

El uso de estudios caso-control para evaluar programas de cribaje cuestiona algunos aspectos metodológicos especiales. Las pruebas realizadas con fines de diagnóstico en lugar de cribaje no deben considerarse y, para los controles, la historia de cribaje debe limitarse hasta el momento del diagnóstico del caso (como en el Ejemplo 16.10) para asegurar que los casos y los controles sean totalmente comparables respecto al periodo de exposición al cribaje.

Los estudios caso-control pueden realizarse como una parte integrante de programas de cribaje en curso para evaluar la estrategia óptima de cribaje (p.e. edad en la que debe iniciarse o pararse el cribaje, o la frecuencia óptima de cribaje).

16.3.3 Grupos de alto riesgo

Una manera de reducir los costes de un programa de cribaje es dirigiendo el programa a grupos de individuos que presentan un riesgo mayor que la media de desarrollar la enfermedad de interés. La mayoría de los programas de cribaje de cáncer están limitados a ciertos grupos de edad. Por ejemplo, no vale la pena cribar para cáncer de mama a mujeres menores de 40 años porque a estas edades se producen muy pocos casos.

Los programas de cribaje también pueden dirigirse exclusivamente a grupos definidos de alto riesgo basándose en la exposición a factores tales como la historia familiar, la historia médica o la ocupación. Por ejemplo, di-

rigiendo programas de cribaje de cáncer de mama exclusivamente a mujeres con una historia familiar positiva, aumentaría la proporción de casos detectados en las mujeres cribadas, pero el programa perdería la gran mayoría de casos en la población porque éstos se producen en personas sin una historia familiar. Por tanto, la restricción de un programa de cribaje a grupos elegidos de alto riesgo sólo es útil si una fracción considerable de los casos totales en una población se produce en estos grupos de alto riesgo.

16.4 Prevencción terciaria

La prevencción terciaria consiste en la reducción de la incapacidad resultante de la enfermedad para mejorar el desenlace final de la enfermedad entre los individuos afectados. No sólo incluye el tratamiento de por sí, sino también todos los intentos de rehabilitación para que el individuo afectado recupere un papel útil, gratificante y, si es posible, autosuficiente en la sociedad.

Los ensayos clínicos aleatorizados son el único método aceptable para evaluar los tratamientos de cáncer (véase Capítulo 7). No obstante, los datos de registros poblacionales de cáncer pueden proporcionar una idea más representativa para la evaluación integral de la atención del cáncer en una determinada población, ya que incluirán todos los casos de cáncer en la población independientemente del tratamiento que puedan o no haber recibido. Estos aspectos se comentarán en el Apartado 17.6.2.

Cuadro 16.4. Puntos esenciales

- La epidemiología es una disciplina clave de salud pública que proporciona la base científica para la formulación de programas destinados a prevenir el desarrollo de la enfermedad en personas sanas.
- Existen tres niveles de prevencción:
 - (a) *Prevencción primaria*, que tiene como objetivo evitar la aparición de la enfermedad mediante una reducción de la exposición a los factores de riesgo o mediante un aumento de la resistencia de los individuos a los mismos. Las medidas de impacto poblacional son muy útiles para identificar las exposiciones potencialmente responsables de grandes cantidades de casos de un determinado cáncer en la población.
 - (b) *Prevencción secundaria*, que corresponde a la detección y tratamiento precoces. Los programas de cribaje son una parte importante de la prevencción secundaria.
 - (c) *Prevencción terciaria*, que tiene como objetivo mejorar el pronóstico y la calidad de vida de los individuos afectados, ofreciéndoles los mejores tratamientos y programas de rehabilitación disponibles.

Lecturas adicionales

* En Cole y Morrison (1978) se presenta una discusión extensa sobre los programas de cribaje de cáncer.

Cuadro 16.4. Puntos esenciales (Cont.)

- * El *cribaje* implica el uso de una prueba simple y barata para detectar el cáncer en una fase inicial cuando el tratamiento es más eficaz que en el momento del diagnóstico y tratamiento habitual. Los programas de cribaje masivo sólo deben dirigirse al control de cánceres para los cuales existe un tratamiento eficaz que reducirá la mortalidad, si se aplican en fases iniciales. También deben ser válidos, económicos y aceptables para la detección del cáncer en fases iniciales.
- La realización de una prueba de cribaje en términos de su aceptabilidad, viabilidad y costes puede monitorizarse mediante *medidas del proceso* relacionadas con los aspectos administrativos y de organización del programa (p.e. proporción examinada de la población diana, funcionamiento del sistema de referimiento, coste por caso detectado, etc.). El valor predictivo de una prueba positiva es una medida especialmente útil porque proporciona una idea de si la mayor parte del esfuerzo del programa se está utilizando para identificar casos en una fase precoz o si se está desperdiciando en la evaluación de falsos positivos. Aunque las medidas del proceso son útiles para monitorizar la actividad del programa *no* indican si los individuos cribados tendrán una mortalidad menor que los no cribados.
- La medida principal de la evaluación de la *efectividad* de un programa de cribaje destinado a la detección precoz del cáncer es la reducción de la mortalidad. Cuando el programa está destinado a detectar las condiciones precancerosas y el cáncer inicial (p.e. frotis Papanicolaou), la reducción en la incidencia de cáncer invasivo y la reducción de la mortalidad son medidas adecuadas. También se han utilizado *medidas intermedias del resultado* tales como la distribución por estadíos y la tasa de letalidad (supervivencia), pero, aunque proporcionan una idea de la posible efectividad del programa, están sujetas al sesgo de duración, al sesgo de adelanto diagnóstico y al sesgo de sobrediagnóstico.
- Idealmente, la efectividad de un programa de cribaje debe evaluarse realizando un ensayo de intervención aleatorizado. Sin embargo, la mayor parte de las evidencias sobre los efectos de los programas de cribaje proceden de estudios observacionales.

Anexo 16.1

Cálculo de las medidas absolutas del efecto de la exposición y de las medidas de impacto poblacional en estudios caso-control

En *estudios caso-control poblacionales* en los cuales se conoce la tasa de incidencia en la población general de interés y en los que la distribución de la exposición en los controles se asume es representativa de toda la población, pueden utilizarse estos parámetros para estimar las tasas de incidencia en los grupos de expuestos y no expuestos.

Una población está formada por una mezcla de personas expuestas y no expuestas. Por tanto, la tasa de incidencia global (r) de la enfermedad en una población es igual a la media ponderada de las tasas de incidencia en los grupos de expuestos (r_1) y no expuestos (r_0), siendo los pesos las proporciones de individuos en cada grupo. Supongamos que una proporción p_e de la población está expuesta al factor en estudio. Por tanto, la proporción de personas no expuestas en esta población es igual a $(1 - p_e)$. Por consiguiente, la tasa en la población será

$$r = r_1 p_e + r_0 (1 - p_e)$$

Debido a que el riesgo relativo (estimado por la *odds ratio* (OR) en estudios caso-control) es la razón de las tasas de incidencia entre los expuestos y los no expuestos, la tasa de incidencia entre los miembros de una población expuestos (r_1) es igual al riesgo relativo por la tasa en los no expuestos ($OR \times r_0$). Por tanto,

$$r = (r_0 \times OR \times p_e) + r_0 (1 - p_e)$$

$$= r_0 [(OR \times p_e) + (1 - p_e)]$$

$$r_0 = \frac{r}{(OR \times p_e) + (1 - p_e)}$$

Una vez se ha determinado la tasa de incidencia en los no expuestos, ésta puede multiplicarse por la *odds ratio* para proporcionar una estimación de la incidencia en los expuestos. Considerando estas dos tasas de incidencia (r_1 y r_0), pueden calcularse el *riesgo en exceso* y la *fracción en exceso*.

Ejemplo A16.1. En un estudio caso-control poblacional realizado en Londres, los casos con cáncer de pulmón presentaban una probabilidad 9 veces mayor de haber fumado regularmente en los últimos 5 años que los hombres sin cáncer de pulmón. La tasa de incidencia de cáncer de pulmón en la población fue de 40 por 100.000 p-as. La proporción de fumadores entre los controles fue del 60%. Así,

$$r_0 = \frac{40 \text{ por } 100.000 \text{ p-as}}{(9 \times 0,6) + (1 - 0,6)} = 6,9 \text{ por } 100.000 \text{ p-as.}$$

$$r_1 = 9 \times 6,9 \text{ por } 100.000 \text{ p-as} = 62,1 \text{ por } 100.000 \text{ p-as.}$$

$$\text{Riesgo en exceso} = 62,1 \text{ por } 100.000 \text{ p-as.} - 6,9 \text{ por } 100.000 \text{ p-as} = 55,2 \text{ por } 100.000 \text{ p-as.}$$

$$\text{Fracción de exceso (\%)} = 100 \times (55,2 \text{ por } 100.000 \text{ p-as} / 62,1 \text{ por } 100.000 \text{ p-as}) = 89\%$$

Obsérvese que la fracción de exceso también se podría haber calculado utilizando la fórmula indicada en el Apartado 16.2.1:

$$\text{Fracción de exceso (\%)} = 100 \times \frac{(OR - 1)}{(OR)}$$

$$\text{Fracción de exceso (\%)} = 100 \times \frac{(9 - 1)}{8} = 89\%$$

El riesgo en exceso de la población y la fracción de exceso de la población se pueden determinar como:

$$\text{Riesgo en exceso de la población} = \text{Riesgo en exceso} \times \text{proporción de la población expuesta al factor (p}_e\text{)}$$

$$\begin{aligned} \text{Fracción de exceso de la población (\%)} &= \\ &= 100 \times \frac{\text{Riesgo en exceso de la población}}{\text{Tasa en la población total (r)}} \end{aligned}$$

La fracción de exceso de la población también puede calcularse utilizando la fórmula indicada en el Apartado 16.2.1:

$$\begin{aligned} \text{Fracción de exceso de la población (\%)} &= \\ &= 100 \times \frac{p_e(OR - 1)}{p_e(OR - 1) + 1} \end{aligned}$$

Ejemplo A16.2. En el ejemplo anterior, el riesgo en exceso de la población y la fracción de exceso de la población puede calcularse como sigue:

Riesgo en exceso de la población = 55,2 por 100.000 p-as. $\times 0,6 = 33,1$ por 100.000 p-as.

$$\text{Fracción de exceso de la población (\%)} = 100 \times \frac{0,6(9-1)}{0,6(9-1)+1} = 83\%$$

Por tanto, un 83% de los casos de cáncer de pulmón en Londres podrían atribuirse al fumar.

En estudios caso-control hospitalarios, no puede calcularse el riesgo en exceso o el riesgo en exceso de la población ya que no pueden estimarse las tasas de incidencia en los expuestos y en los no expuestos. No obstante, para calcular la fracción de exceso y la fracción de exceso de la población pueden utilizarse las fórmulas siguientes

$$\text{Fracción de exceso (\%)} = 100 \times \frac{(OR-1)}{(OR)}$$

$$\text{Fracción de exceso de la población (\%)} = 100 \times \frac{p_e(OR-1)}{p_e(OR-1)+1}$$

Capítulo 17

El papel de los registros de cáncer

17.1 Objetivos de los registros de cáncer

El registro de cáncer organiza la recopilación, almacenamiento, análisis, interpretación y notificación sistemáticas de datos sobre personas con cáncer. Existen dos tipos principales de registros de cáncer: hospitalarios y poblacionales.

Los *registros de cáncer hospitalarios* registran la información sobre los pacientes con cáncer diagnosticados y/o tratados en un hospital particular. El principal objetivo de estos registros es contribuir al cuidado del paciente, facilitando el acceso a la información sobre los sujetos visitados con cáncer, el tratamiento recibido y su resultado. Los datos se utilizan principalmente con fines administrativos y control de la eficacia clínica. Aunque estos datos pueden utilizarse, en cierto grado, con fines epidemiológicos (véase Apartado 17.7), estos registros raramente pueden proporcionar medidas de la incidencia de cáncer en una población definida ya que es difícil establecer sus poblaciones de captura, es decir las poblaciones de las que proceden los casos.

Los *registros de cáncer poblacionales* recopilan datos sobre todos los casos nuevos de cáncer que se producen en una población bien definida. Normalmente, la población es la residente en una región geográfica particular. Como resultado, y a diferencia de los registros hospitalarios, el principal objetivo es producir estadísticas sobre la incidencia de cáncer en una población definida y proporcionar un marco para la evaluación y control del impacto del cáncer en la comunidad. Por tanto, son importantes en el campo de la epidemiología y salud pública.

El uso de los registros de cáncer poblacionales puede resumirse en:

- (1) Describen la magnitud y naturaleza del impacto del cáncer en la comunidad y ayudan en el establecimiento de prioridades de salud pública.
- (2) Pueden utilizarse como fuente de material para estudios etiológicos.
- (3) Ayudan a monitorizar y evaluar la eficacia de las actividades en el control del cáncer.

Algunas de estas funciones pueden realizarse mediante los datos de mortalidad obtenidos por los sistemas de estadísticas vitales. No obstante, los datos de los registros de cáncer proporcionan una información más exten-

sa, válida y detallada sobre las características del paciente de la que puede obtenerse a partir de los certificados de defunción. Por otro lado, en la mayoría de los países desarrollados, pero en sólo unos pocos países subdesarrollados, existen datos de mortalidad fiables. Por tanto, los registros de cáncer pueden ser la única fuente de información sobre la magnitud y los patrones del cáncer en estos países y ayudar a la investigación tanto en la etiología como en la prevención.

Si no se indica lo contrario, la discusión en el resto de este capítulo se centrará en registros del cáncer de base poblacional.

17.2 Breve historia de los registros de cáncer

Los primeros intentos serios de cuantificar el cáncer se realizaron a principios de siglo en varios países europeos. En Alemania, el 15 de octubre de 1900 (Anon., 1901), se intentó registrar a todos los pacientes de cáncer en tratamiento médico mediante cuestionarios remitidos a cada médico del país. Este método se repitió entre 1902 y 1908 en Dinamarca, España, Holanda, Hungría, Islandia, Portugal y Suecia. No obstante, estos esfuerzos no tuvieron éxito, principalmente debido a la baja colaboración de los médicos. Se realizaron estudios transversales similares en los Estados Unidos de América.

El primer registro de cáncer poblacional se realizó en Hamburgo (Alemania) en 1926. Tres enfermeras visitaron los hospitales y los médicos de la ciudad a intervalos regulares. Registraron los nombres de los pacientes de cáncer nuevos y transfirieron los datos a un índice central en el departamento de sanidad. Este índice se comparó una vez a la semana con los certificados de defunción oficiales. En las décadas subsiguientes se realizaron otros registros de cáncer poblacionales de modo que, hacia 1955, ya se habían implantado casi veinte registros poblacionales en varios países (Tabla 17.1).

Actualmente, existen más de 200 registros de cáncer poblacionales en diferentes partes del mundo. Abarcan aproximadamente un 5% de la población mundial, pero esta proporción es mucho mayor en países desarrollados que en los subdesarrollados. Además, en los países en vías de desarrollo, es más frecuente que los registros cubran áreas urbanas donde la accesibilidad a los servicios de diagnóstico y tratamiento es más fácil.

Los registros de cáncer nacionales operan en algunos países tales como Inglaterra y Gales, Escocia, los países nórdicos, Canadá, Australia, Nueva Zelanda, Israel, Cuba, Puerto Rico y Gambia. El registro de cáncer danés, fundado en 1942, es el registro más antiguo en funcionamiento que incluye una población nacional. No obstante, en la mayoría de los países, los registros de cáncer poblacionales cubren sólo una proporción de la población (p.e. India, Colombia, Estados Unidos, Italia). También se han realizado algunos registros especializados que cubren sólo grupos de edad específicos (p.e. cánceres infantiles en Oxford, Inglaterra) o tipos de cáncer concretos (p.e. cánceres gastrointestinales en Dijon, Francia). Además, en numerosos hospitales de todo el mundo se han iniciado registros de cáncer hospitalarios.

La Asociación Internacional de Registros de Cáncer (IACR) se formó en 1966. El principal objetivo de esta asociación es desarrollar y estandarizar el contenido y los métodos de recopilación de datos sobre pacientes de cáncer mediante registros con el fin de presentar los datos de la forma más comparable posible.

Una explicación más detallada de la historia de los registros de cáncer se encuentra en Wagner (1991).

País (región)	Año de realización	Notificación
RF Alemania (Hamburgo)	1929	Voluntaria
EE.UU. (Estado de Nueva York)	1940	Obligatoria
EE.UU. (Connecticut)	1941 (casos registrados retrospectivamente desde 1935)	Obligatoria (desde 1971)
Dinamarca	1942	Obligatoria (desde 1987)
Canadá (Saskatchewan)	1944	Obligatoria
Inglaterra y Gales (Región SO)	1945	Voluntaria
Inglaterra y Gales (Liverpool)	1948	Voluntaria
Nueva Zelanda	1948	Obligatoria
Canadá (Manitoba)	1950	Voluntaria
Eslovenia	1950	Obligatoria
Canadá (Alberta)	1951	Obligatoria
EE.UU. (El Paso)	1951	Voluntaria
Hungría (Szabolcs, Miskolc, Vas)	1952	Obligatoria
Noruega	1952	Obligatoria
Antigua USSR	1953	Obligatoria
Antigua República Democrática Alemana	1953	Obligatoria
Finlandia	1953	Obligatoria (desde 1961)
Islandia	1954	Voluntaria

^a Reproducido con la autorización de Wagner (1991).

Tabla 17.1.
Registros de cáncer poblacionales establecidos antes de 1955^a.

17.3 Metodología de un registro de cáncer

El objetivo de un registro de cáncer poblacional es recopilar información sobre cada caso de cáncer identificado dentro de una población específica en un periodo de tiempo determinado. Para asegurar esto, antes de iniciar un registro de cáncer poblacional es necesario garantizar que se cumplan los siguientes requisitos básicos:

- (a) Definir claramente la población a estudiar. El registro debe poder distinguir entre los residentes de la zona y los procedentes de fuera y debe poder registrar los casos en personas residentes que son tratados fuera de la zona geográfica definida.
- (b) Disponibilidad de denominadores fiables de la población a partir del censo u otras oficinas estadísticas.
- (c) Atención médica disponible y acceso directo de la población general a las instalaciones médicas, de modo que la mayoría de los casos de cán-

cer entrarán en contacto con el sistema sanitario en algún momento de su enfermedad y, por tanto, se diagnosticarán correctamente.

(d) Acceso fácil a las fuentes de búsqueda de casos tales como los hospitales, departamentos de anatomía patológica, certificados de defunción y otras fuentes de datos clínicos dentro de la zona geográfica definida y de las zonas que la rodean.

17.3.1 Recopilación de datos

La forma en que funciona un registro depende estrechamente de las condiciones locales y de las fuentes de material disponibles. Normalmente, las principales fuentes de información de un registro poblacional incluyen: (1) información de los centros de tratamiento, tales como los centros de cáncer y los principales hospitales (y a veces, si procede, clínicas privadas, hospicios, hogares de ancianos y médicos generales); (2) información de los servicios de diagnóstico, especialmente departamentos de anatomía patológica, laboratorios de hematología, bioquímica e inmunología y servicios de radiología; (3) certificados de defunción (si están disponibles).

La información de estas fuentes puede obtenerse de dos formas. La *recopilación activa* implica que el personal del registro visite las diferentes fuentes y obtenga los datos en formularios especiales. Este es el método habitual de los registros en los países subdesarrollados. La *notificación pasiva* implica que el personal sanitario rellene los formularios de notificación diseñados y distribuidos por el registro o envíe copias al registro de los informes de alta. En la mayoría de los registros en países desarrollados se sigue una combinación de los dos procedimientos, con especial interés en el último. En algunos países, la notificación de casos de cáncer es obligatoria. No obstante, esto no asegura necesariamente la exhaustividad.

Los datos recopilados por un registro vienen dictados por el objetivo para el que se instauró el registro, por el método de recopilación de datos utilizado y por los recursos disponibles para el registro. No obstante, debe prestarse especial atención a la *calidad de los datos obtenidos en lugar de la cantidad*. Se recomienda que los registros en países en desarrollo empiecen intentando obtener sólo información sobre los aspectos básicos indicados en la Tabla 17.2.

El registro asigna a cada paciente un *número de registro* único (número del registro de cáncer). Si un paciente presenta más de un tumor primario, se asigna el mismo número a cada tumor. Los tumores primarios múltiples se distinguirán basándose en su fecha de incidencia, en su topografía y en su morfología.

Otros aspectos de identificación tales como nombre, sexo y fecha de nacimiento (o edad aproximada, si no se conoce la fecha de nacimiento) son importantes para evitar registros múltiples del mismo paciente o tumor, para obtener datos de seguimiento y poder realizar cualquier cruzamiento con otros registros. La dirección habitual del paciente es esencial para conocer la zona de residencia, excluir todos los pacientes no residentes, reali-

Variable	Comentarios
El paciente	
<i>Identificación personal</i>	
Número de registro	Asignado por el registro
Nombre	Según la práctica local
Sexo	
Fecha de nacimiento o edad	Estimación si no se conoce
<i>Demográfico</i>	
Dirección	Residencia habitual
Grupo étnico	Si procede
Tumor	
Fecha de incidencia	
Base de diagnóstico más válida	No microscópica o microscópica
Topografía (localización)	Codificado mediante CIE-O ^b
Morfología (histología)	Codificado mediante CIE-O
Comportamiento	Codificado mediante CIE-O
Fuente de información	Tipo de fuente: médico, laboratorio, hospital, certificado de defunción u otros Fuente real: nombre del médico, laboratorio, hospital, etc. Fechas (p.e. fechas de citas relevantes, ingresos en el hospital, procedimientos de diagnóstico)
^a Modificado de MacLennan (1991).	
^b Clasificación Internacional de Enfermedades para Oncología (Percy et al., 1990).	

Tabla 17.2.

Datos básicos a recopilar por un registro de cáncer poblacional^a.

zar el análisis según zonas de residencia y para el seguimiento de los pacientes. Los datos sobre el carácter étnico son importantes en poblaciones donde hay grupos étnicos diferenciados.

La *fecha de incidencia* es la fecha de la primera consulta o admisión a un hospital/clínica por cáncer, ya que es un punto definido en el tiempo, consistente, fiable y que puede comprobarse en archivos. Esta fecha se elige como la fecha del diagnóstico para los cálculos de incidencia y como la fecha de inicio para el análisis de supervivencia (véase Apartado 17.6.2.). Si no se dispone de esta información, la fecha de incidencia elegida será la fecha del primer diagnóstico de un médico o la fecha del primer informe anatómo-patológico. Un problema especial surge cuando el cáncer se detecta por primera vez en un certificado de defunción y los intentos de seguirlo retrospectivamente no funcionan. En estos casos la fecha de fallecimiento debe tomarse como la fecha de incidencia.

Para evaluar las tasas de incidencia es muy importante conocer *en qué se ha basado el diagnóstico* (cual es la prueba diagnóstica más válida). El registro debe discriminar entre los tumores diagnosticados microscópicamente y los que no. Si es posible, debe recogerse información para poder discernir entre aquellos tumores diagnosticados únicamente por anamnesis, de los que se ha realizado además alguna exploración clínica (como Rayos X), o de los que se ha realizado cirugía exploratoria, histología y/o citología o autopsia. Para poder verificar este punto es importante que el registro recoja información sobre las

Tabla 17.3.

Otras variables opcionales de información que pueden recopilarse en los registros de cáncer poblacionales^a.

<p>Paciente</p> <p><i>Identificación</i></p> <p>Número de identificación personal (p.e. número nacional de identidad o número de la seguridad social)</p> <p><i>Aspectos demográficos y culturales</i></p> <p>Lugar de nacimiento</p> <p>Estado civil</p> <p>Edad en la fecha de incidencia</p> <p>Nacionalidad</p> <p>Religión</p> <p>Ocupación e industria</p> <p>Año de inmigración</p> <p>País de nacimiento del padre y/o madre</p> <p>Tumor y sus investigaciones</p> <p>Certeza del diagnóstico</p> <p>Método de primera detección</p> <p>Extensión clínica de la enfermedad antes del tratamiento</p> <p>Extensión quirúrgica-patológica de la enfermedad antes del tratamiento</p> <p>Sistema TNM</p> <p>Localizaciones de metástasis distantes</p> <p>Tumores primarios múltiples</p> <p>Lateralidad</p> <p>Tratamiento</p> <p>Tratamiento inicial</p> <p>Seguimiento</p> <p>Fecha de último contacto</p> <p>Estado en el último contacto (vivo, muerto, emigrado, desconocido)</p> <p>Fecha de fallecimiento</p> <p>Causa de fallecimiento</p> <p>Lugar de fallecimiento</p> <hr/> <p>^a Modificado de MacLennan (1991).</p>

fuentes de casos (p.e. nombre de médico, hospital, laboratorio), fechas relevantes de acontecimientos médicos (p.e. admisión al hospital, biopsia) y cualquier otro detalle que ayude a trazar la historia clínica del paciente (p.e. número de hospital, número de biopsia, número de referencia del laboratorio).

La inclusión de otros datos además de los indicados en la Tabla 17.2, aumenta la complejidad y los costes de registro y, por consiguiente, sólo debe realizarse si está justificado por las necesidades locales y si se dispone de los recursos necesarios. En la Tabla 17.3 se presenta una lista de las variables opcionales; las más relevantes son la extensión clínica de la enfermedad antes del tratamiento (fase de diagnóstico) y los datos de seguimiento.

Los datos de las diferentes fuentes de casos se suelen obtener utilizando un formulario de registro estándar desarrollado según las necesidades del registro. Al desarrollar un formulario de registro deben tenerse en cuenta dos consideraciones principales:

- (1) La información sobre los casos de cáncer se debe recoger y clasificar de tal modo que se corresponda con los datos disponibles del censo u otras estadísticas. Esto es fundamental para asegurar la comparabilidad entre los numeradores (es decir, el número de cánceres registrados) y los denominadores (es decir, los datos de la población) para el cálculo de las tasas de incidencia.
- (2) Aunque los datos deben recopilarse (y notificarse) según las necesidades y los intereses locales, se debe intentar asegurar que sean comparables con los datos de otros registros de cáncer nacionales e internacionales.

17.3.2 Clasificación y codificación de las neoplasias

Como se menciona en el Anexo 2.2, se recomienda que los registros de cáncer utilicen la *Clasificación Internacional de Enfermedades para Oncología* (CIE-O) (Percy *et al.*, 1990) para codificar la topografía (localización del tumor primario) y la morfología (tipo histológico) de los tumores. El quinto dígito en los códigos CIE-O de la morfología describe el comportamiento del tumor –benigno, incierto si benigno o maligno, *in-situ*, maligno. La topografía de un tumor es el dato registrado más importante y proporciona la base principal de tabulación de los datos.

17.3.3 Calidad de los datos

Deben considerarse dos aspectos principales al evaluar la calidad de los datos de un registro de cáncer: su *exhaustividad* y su *validez*. Un registro poblacional debe, por definición, registrar todos y cada uno de los casos que se producen en la población definida. No obstante, la recogida de casos raramente es completa. Se utilizan varios métodos para determinar el grado de exhaustividad del registro tales como comparaciones con certificados de defunción y archivos de hospitales. También es importante determinar la capacidad del registro para eliminar los casos de personas residentes fuera de la población de captura y evitar registros múltiples de la misma persona o del mismo tumor.

La validez de los datos puede evaluarse de varias formas. La proporción de casos con confirmación histológica del diagnóstico es un índice muy útil, así como es la proporción de casos detectados en vida (no simplemente a partir de un certificado de defunción). Los registros de cáncer deben desarrollar sus propios controles de calidad internos de modo que se preste atención a la información que falta y a la presencia de datos incongruentes. Con frecuencia, muchos registros vuelven a reunir y codificar una muestra de casos para evaluar la calidad de sus datos. Parkin *et al.* (1994) presenta una discusión completa de los métodos de control de calidad de un registro.

17.3.4 Notificación de los resultados

La recopilación de información sobre casos de cáncer y la producción de estadísticas de cáncer sólo están justificadas si los datos recopilados se utilizan.

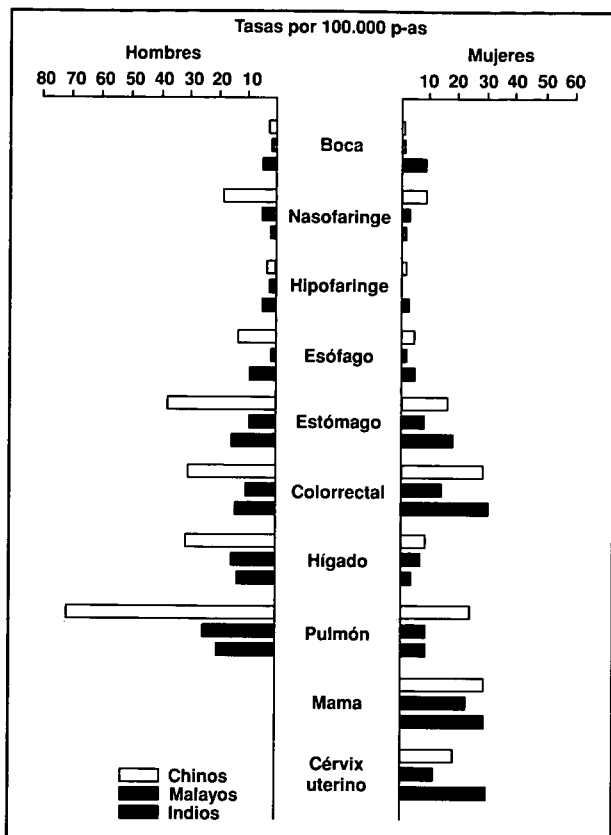


Figura 17.1. Tasas de incidencia estandarizadas por edad (para la población mundial) para algunas localizaciones tumorales según sexo y grupo étnico, Singapur, 1978-1982 (reproducido con la autorización de Lee *et al.*, 1988).

Un registro de cáncer poblacional debe hacer disponibles sus datos y resultados en forma de informes y artículos en revistas científicas. Los informes deben incluir la información propia del registro, los procedimientos de registro, la población de captura, el grado de exhaustividad de los datos y la validez, los métodos de análisis y los resultados. Deben producirse estadísticas básicas y presentar los datos, principalmente según la topografía del tumor. Los datos deben presentarse en forma tabulada o como gráficos. En la Figura 17.1 y Tabla 17.4 se presentan unos ejemplos.

17.4 Registro de cáncer en países en desarrollo

El registro de cáncer podría aparecer como una prioridad secundaria de los servicios sanitarios de un país en desarrollo, dadas todas las demandas que compiten por los recursos, normalmente escasos, asignados a la salud. No obstante, el cáncer ya es un problema sanitario significativo en muchos de estos países. Más de la mitad de los casos nuevos de cáncer en el mundo aparecen en ellos (Parkin *et al.*, 1993). El rápido aumento en la esperanza de vida (debido a una reducción en la mortalidad por enfermedades infecciosas) junto con la adopción de estilos

de vida occidentales parece anticipar que el impacto del cáncer en estos países tenderá a aumentar en un futuro próximo.

Con frecuencia, los registros de cáncer proporcionan la única oportunidad de evaluar correctamente la magnitud y naturaleza de la magnitud del cáncer en estos países, ya que muy pocos de ellos poseen datos de mortalidad fiables. Idealmente, el objetivo es establecer un registro de cáncer poblacional que permita determinar la incidencia de diferentes tumores en una comunidad bien definida. En la práctica, los registros de cáncer en países en desarrollo con frecuencia empiezan basándose en los casos que acuden a ciertos hospitales o en departamentos de histopatología, debido a la relativa facilidad con que este tipo de registros pueden organizarse.

Los registros de cáncer poblacionales en países subdesarrollados normalmente presentan enormes problemas logísticos debido a la falta de personal correctamente formado y de recursos adecuados. Finalmente, la organización de un registro puede comprometerse por factores externos de difícil control.

Falta de servicios sanitarios básicos

El papel del registro de cáncer depende en gran parte de la disponibilidad de servicios sanitarios adecuados para el diagnóstico y tratamiento de pacientes oncológicos. En muchos países en desarrollo, las instalaciones sanitarias son escasas y tienden a estar concentradas en zonas urbanas y la ca-

Localización (CIE-9)	Número de casos por grupo de edad									Total	%	Tasa de incidencia	
	Desconocido	0-	15-	25-	35-	45-	55-	65-	75+			Cruda	TEE ^b
Todas las localizaciones	10	69	89	255	241	266	362	264	74	1630	100.0	101.1	238.5
Todas las localizaciones excepto piel	8	69	88	253	236	264	359	259	72	1608		99.8	234.6
Cavidad oral (140-145)	1	-	1	1	2	5	2	3	1	16	1.0	1.0	2.5
Nasofaringe (147)	-	-	1	5	1	1	1	4	-	13	0.8	0.8	2.0
Otros faringe (148-149)	-	-	1	-	-	-	2	1	-	4	0.2	0.2	0.6
Esófago (150)	-	-	-	1	16	25	63	35	13	153	9.4	9.5	30.4
Estómago (151)	-	-	-	2	10	15	14	20	6	67	4.1	4.2	13.5
Colon (153)	-	-	5	2	6	6	9	9	2	39	2.4	2.4	6.6
Recto (154)	-	-	2	3	6	8	4	4	1	28	1.7	1.7	3.8
Hígado (155)	-	2	10	22	37	41	46	46	9	213	13.1	13.2	34.6
Páncreas (157)	-	-	-	1	2	7	13	7	1	31	1.9	1.9	5.9
Laringe (161)	1	-	-	-	-	4	12	4	2	23	1.4	1.4	4.5
Bronquios y pulmón (162)	-	-	-	1	6	30	50	32	6	125	7.7	7.8	24.6
Pleura (163)	-	-	-	-	-	-	1	-	-	1	0.1	0.1	0.1
Tejido conectivo (171)	-	4	4	4	2	2	1	-	-	17	1.0	1.1	1.1
Melanoma cutáneo (172)	-	-	2	2	2	2	4	1	1	14	0.9	0.9	1.8
Otros piel (173)	2	-	1	2	5	2	3	5	2	22	1.3	1.4	4.0
Mama (175)	-	-	-	-	1	3	1	-	-	5	0.3	0.3	0.6
Próstata (185)	3	-	-	-	2	11	37	41	18	112	6.9	6.9	29.2
Pene (187)	-	-	-	-	1	4	3	2	3	13	0.8	0.8	2.8
Vejiga urinaria (188)	-	1	-	3	5	19	18	16	6	68	4.2	4.2	13.2
Riñón (189)	-	10	1	-	1	2	1	2	-	17	1.0	1.1	1.7
Ojo (190)	-	5	1	1	-	1	1	1	-	10	0.6	0.6	0.9
Encéfalo,													
Sistema nervioso (191-192)	-	7	6	4	5	2	4	1	-	29	1.8	1.8	2.4
Tiroides (193)	-	-	1	1	1	1	4	1	-	9	0.6	0.6	1.2
Enfermedad de Hodgkin (201)	-	2	1	4	2	2	2	-	-	13	0.8	0.8	1.0
Linfoma no Hodgkin (200, 202)	-	12	4	13	11	10	6	2	-	58	3.6	3.6	4.7
Mieloma múltiple (203)	-	-	-	1	5	4	8	1	1	20	1.2	1.2	2.7
Leucemia linfocítica (204)	-	8	3	1	1	1	2	4	-	20	1.2	1.2	2.5
Leucemia mieloide (205)	-	8	6	6	5	6	2	1	-	34	2.1	2.1	2.7
Otras leucemias (207-208)	-	-	-	-	-	-	1	1	-	2	0.1	0.1	0.6
Sarcoma de Kaposi	2	7	28	171	97	44	27	4	-	380	23.3	23.6	24.6
Otros e inciertos	1	3	11	4	9	8	20	16	2	74	4.5		

^a Reproducido con permiso de Wiley-Liss Inc., subsidiaria de John Wiley & Sons Inc., de Bassett et al. (1995).

^b TEE = Tasa de incidencia estandarizada por edad a la población mundial.

alidad de la información del diagnóstico puede ser limitada, basada únicamente en el examen clínico.

Falta de denominadores adecuados

Los registros poblacionales requieren información sobre el tamaño y la naturaleza de la población de referencia, información que precisa de la disponibilidad del censo. Los censos son especialmente difíciles de realizar en países en desarrollo, y por tanto, tienden a realizarse con poca frecuencia y sus resultados suelen estar disponibles con retraso y con información poco detallada.

La población de muchos de estos países es especialmente móvil debido a la tendencia creciente a migrar temporalmente de zonas rurales a urbanas y debido a que las circunstancias sociales y políticas pueden forzar a comunidades enteras a desplazarse de una zona a otra. Por tanto, las estimaciones intercensales o proyecciones postcensales del tamaño y la estructura de la población es probable que sean inexactas.

Tabla 17.4.

Ejemplo del tipo de tabla utilizado por los registros de cáncer para notificar sus datos. Número de registros de cáncer en hombres africanos residentes en Harare, Zimbabwe, 1990-1992. Registro de cáncer de Harare, Zimbabwe ^a.

Estos cambios de la población presentan un reto especial para los registros de cáncer, que deben esforzarse especialmente en distinguir los residentes de los no residentes en su zona de captura utilizando, en lo que sea posible, las mismas definiciones que en el censo.

Identidad de los individuos

La capacidad de distinguir a los individuos es un aspecto clave de un registro de cáncer. Así, el registro debe tener la suficiente información sobre cada individuo para evitar registros múltiples del mismo sujeto. El identificador más universal y utilizado generalmente es el nombre. La utilidad de usar nombres variará dependiendo de las costumbres locales. Por ejemplo, puede no utilizarse el apellido – las personas pueden conocerse sólo por su nombre. Los individuos pueden cambiar de nombre cuando se casan o por otras razones sociales. Las variaciones en la escritura de los nombres es un problema frecuente, especialmente cuando un gran porcentaje de la población es analfabeta. Esto se agrava si existe una necesidad de traducir los nombres al alfabeto romano, con objeto de utilizar sistemas de bases de datos computerizados.

Falta de seguimiento

El seguimiento activo normalmente significa que el registro intenta ponerse en contacto regularmente con médicos o pacientes para ver si el paciente todavía está vivo. Debido a que esto es caro, muchos registros se basan en el seguimiento pasivo; la confrontación con los certificados de defunción y la asunción de que si no constan en el registro de mortalidad los pacientes están vivos. Los sistemas mixtos utilizan certificados de defunción con actualización de la 'última fecha conocida como vivo' a partir de la admisión al hospital, consultas y otras fuentes de datos.

El seguimiento activo de los pacientes normalmente es muy difícil en países en desarrollo. Pocos registros disponen de recursos que permitan seguir regularmente a sus pacientes. También existen problemas con los servicios postales poco fiables, direcciones inestables y la movilidad de la población. El seguimiento pasivo sólo es posible en los países, pocos, en los que existe un sistema fiable de registro de mortalidad.

17.5 El papel de los registros de cáncer en epidemiología

Los registros de cáncer poblacionales son recursos importantes para los epidemiólogos de cáncer ya que ofrecen información sobre la distribución de la enfermedad en poblaciones bien definidas. Esta información puede analizarse sin necesidad de recopilación de datos adicionales. Las *tasas de incidencia* de cáncer por localizaciones tumorales se pueden calcular y comparar según muchas variables diferentes tales como edad, sexo, país de nacimiento, lugar de residencia en el momento del diagnóstico, etc. También pueden realizarse estudios de tendencias temporales cuando se han acu-

Ejemplo 17.1. El Estudio de Intervención de Hepatitis de Gambia [The Gambia Hepatitis Intervention Study] es un ensayo de vacunación a gran escala en Gambia, iniciado en julio de 1986, en el que aproximadamente 60.000 niños recibieron tres dosis de vacuna antihepatitis B y un número similar no la recibió. En el registro nacional de cáncer se determinarán los casos nuevos de cáncer de hígado (The Gambia Hepatitis Study Group, 1987).

Ejemplo 17.2. La importancia de algunos factores de riesgo en la etiología del cáncer de esófago en Bulawayo, Zimbabwe, se examinó utilizando datos recopilados por el registro de cáncer local durante los años 1963-1977, cuando se intentó entrevistar a todos los pacientes mediante un cuestionario estandarizado. Los factores de riesgo para el cáncer de esófago se estimaron por análisis caso-control en el que se tomaron como grupo 'control' otros cánceres no relacionados con tabaco y alcohol. La Tabla 17.5 muestra el análisis en varones. Se observó una fuerte asociación con el consumo de tabaco, con un efecto aparente dosis-respuesta. Por el contrario, el consumo de alcohol pareció tener poco efecto sobre el riesgo de cáncer de esófago en esta población (Vizcaino et al., 1995).

	Casos	Controles ^b	Odds ratio (intervalo de confianza al 95%) ^c
Consumo de tabaco			
No fumador ^d	120	947	1,0
Ex fumador	21	38	3,4** (1,9-6,2)
< 15 g diarios	279	542	3,5** (2,7-4,5)
≥ 15 g diarios	71	91	5,7** (3,8-8,4)
No especificado	56	116	2,8** (1,8-4,2)
<i>Test de tendencia</i>			<i>P < 0,001</i>
Consumo de alcohol			
Ninguno ^d	144	654	1,0
Ocasionalmente	44	206	0,6* (0,4-0,9)
Semanal	121	387	0,8 (0,6-1,1)
Diario	212	539	0,9 (0,7-1,2)
No especificado	41	68	1,8* (1,1-3,0)

^a Datos de Vizcaino et al. (1995)

^b Formado por todos los otros cánceres no relacionados con tabaco y alcohol (es decir, tras la exclusión de cánceres de la cavidad oral y faringe, hígado, laringe, pulmón y vejiga).

^c Ajustado por la edad, provincia de residencia, ocupación y según las otras variables de la tabla.

^d Categoría basal

* $P < 0,05$; ** $P < 0,001$.

Tabla 17.5. Factores de riesgo de cáncer de esófago en hombres, Sudoeste de Zimbabwe, 1963-1977^a.

mulado datos durante periodos prolongados de tiempo. Los métodos utilizados en este tipo de análisis se comentaron en los Capítulos 4 y 11. Los datos de los registros de cáncer poblacionales de todo el mundo están publicados en la serie Incidencia del Cáncer en Cinco Continentes [*Cancer Incidence in Five Continents*] (Doll *et al.*, 1966; Waterhouse *et al.* 1970, 1976, 1982; Muir *et al.*, 1987; Parkin *et al.*, 1992, 1997). Estos datos son muy valiosos para las comparaciones internacionales.

Además de los datos de incidencia, los registros de cáncer poblacionales que realizan algún tipo de seguimiento de sus pacientes pueden estimar la *prevalencia* del cáncer. Los datos de prevalencia proporcionan una indicación del peso de esta enfermedad en la comunidad. Los registros de cáncer generalmente asumen que una vez diagnosticado un individuo con cáncer, éste sigue siendo un caso prevalente hasta la muerte. Por tanto, la prevalencia puede estimarse a partir de datos sobre la incidencia y la supervivencia. No obstante, cuando un registro ha estado funcionando durante muchos años, de modo que todos los pacientes diagnosticados con cáncer antes de la realización del registro han muerto, los casos prevalentes pueden enumerarse simplemente a partir del registro siempre que éste reciba información sobre las muertes y las migraciones para todos los casos registrados.

El registro de cáncer proporciona un método económico y eficaz para detectar casos de cáncer en *ensayos de intervención* (Ejemplo 17.1) y *estudios de cohorte*, siempre que los pacientes estén identificados correctamente en los archivos.

Ejemplo 17.3. *Un grupo colaborativo de registros de cáncer poblacionales y centros de tratamiento del cáncer efectuaron un estudio caso-control para identificar las razones de los aumentos observados del riesgo de cáncer de pulmón tras la enfermedad de Hodgkin. Se identificaron un total de 98 casos de cáncer de pulmón en pacientes que habían sobrevivido como mínimo un año después del diagnóstico de enfermedad de Hodgkin. Para cada caso, se eligieron tres controles con enfermedad de Hodgkin que no desarrollaron cáncer de pulmón apareados con el caso respecto al registro u hospital, sexo, año de nacimiento y año de diagnóstico de la enfermedad de Hodgkin. Para casos y controles se obtuvo información detallada a partir de las historias clínicas referentes al estadio y tratamiento de la enfermedad Hodgkin. Los pacientes tratados sólo con quimioterapia presentaron un riesgo de desarrollar cáncer de pulmón de aproximadamente el doble que aquellos tratados sólo con radioterapia o con las dos modalidades (Kaldor *et al.*, 1992).*

Los registros de cáncer basados en la población también pueden convertirse en una fuente de casos para *estudios caso-control*. En general, los registros de cáncer no se consideran tan adecuados para la realización de estos estudios debido a los retrasos propios de los registros en la detección de casos. El principal valor del registro en estos estudios es evaluar la exhaustividad y la representatividad del grupo de casos estudiado.

El registro puede efectuar sus propios estudios caso-control dentro de su base de datos, comparando un tipo de cáncer con una selección de los otros cánceres («controles») (véase Apartado 11.1.6). Las variables disponibles habitualmente para estos análisis están limitadas a la recogida rutinaria del registro. Los registros pueden complementar estas variables con información adicional (p.e. fumar, dieta, ocupaciones, tratamientos, etc.) entrevistando muestras de pacientes (Ejemplo 17.2), extrayendo esta información de archivos médicos o por enlace de archivos.

Los registros de cáncer han sido especialmente útiles en la realización de estudios caso-control para investigar los efectos carcinogénicos de algunos tratamientos de cáncer (Ejemplo 17.3).

17.6 Papel de los registros de cáncer en el control del cáncer

El registro de cáncer es una parte esencial de todo programa de control del cáncer. Sus datos pueden utilizarse en una gran variedad de áreas que abarcan desde la investigación etiológica, pasando por la prevención primaria y secundaria a la planificación sanitaria y el cuidado del paciente. Aunque la mayoría de los registros de cáncer no están obligados más que a proporcionar información de base necesaria para estas actividades, poseen el potencial de desarrollar y avalar programas de investigación utilizando la información que recopilan.

17.6.1 Planificación de programas de control del cáncer

La información exacta sobre la aparición del cáncer es importante para fijar prioridades y marcar objetivos de las actividades de control. Los regis-

Localización	CIE-10	Nº de casos incidentes, 1992 ^b	Nº de casos prevalentes el 31 Diciembre 1992 ^b	Supervivencia relativa a 5 años(%) ^c
Pulmón	C33-34	6.434 (1)	8.201 (4)	9
Próstata	C61	4.096 (2)	13.564 (3)	49
Colorrectal	C18-21	3.492 (3)	14.470 (2)	43
Vejiga	C67	2.183 (4)	16.538 (1)	70
Estómago	C16	1.516 (5)	2.407 (7)	13
Linfoma no Hodgkin	C82-85	1.009 (6)	4.582 (5)	51
Esófago	C15	858 (7)	805 (10)	8
Páncreas	C25	833 (8)	586 (11)	6
Riñón	C64	644 (9)	2.296 (8)	40
Cerebro	C71	523 (10)	1.558 (9)	22
Melanoma cutáneo	C43	367 (11)	2.855 (6)	71
Todas las neoplasias malignas (excepto el cáncer cutáneo no melanoma)		28.732	109.637	36

^a Datos del Registro del Cáncer del Támesis, Inglaterra (1995).
^b El orden de frecuencia decreciente se indica entre paréntesis.
^c Pacientes mayores de 15 años, diagnosticados durante los años 1986-89.

Tabla 17.6. Número (y orden) de casos de cáncer masculino incidente y prevalente y tasa de supervivencia relativa a 5 años. Sudeste de Inglaterra, 1992^a.

tros de cáncer poblacionales se encuentran en una posición única para proporcionar esta información.

El número de casos incidentes proporciona una idea de los recursos necesarios para el primer tratamiento mientras que el número de casos prevalentes describe cuántas personas necesitan un seguimiento regular a largo plazo (aunque para algunos cánceres no es necesario una vigilancia regular después de los primeros 5-10 años del diagnóstico). La Tabla 17.6 muestra el número de casos de cáncer incidentes (nuevos) y prevalentes (nuevos y viejos) en el sudeste de Inglaterra, en 1992. El orden de los tipos de cáncer es bastante diferente para la incidencia y la prevalencia. Esto es debido a diferencias en la supervivencia. Los cánceres con un buen pronóstico tienen prevalencias altas aunque su incidencia sea baja, mientras que aquellos con un mal pronóstico presentan prevalencias menores aunque la incidencia sea más alta.

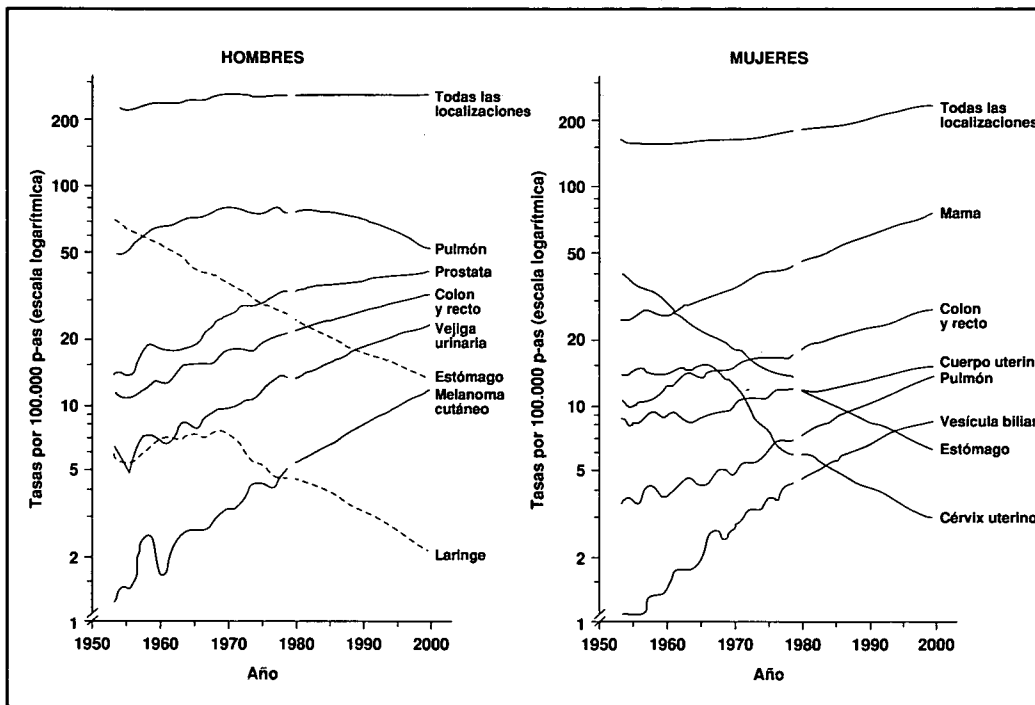


Figura 17.2. Tasas de incidencia anuales ajustadas por edad de algunos cánceres en Finlandia: tasas reales de 1953 a 1979 y predicciones hasta el 2000, basándose en un modelo estadístico que incluye los efectos de la edad, periodo y cohorte (reproducido con la autorización de Läärä, 1982).

te importante de datos para estas predicciones. Las predicciones más simples de las tasas de incidencia de cáncer se basan en continuar las tendencias actuales específicas por edad. La previsión puede mejorarse si también se consideran los efectos de la cohorte de nacimiento utilizando modelos estadísticos de edad-periodo-cohorte (véase el Apartado 4.3.2). En la Figura 17.2 se presenta un ejemplo.

Puede realizarse una estrategia todavía más sofisticada para predecir las tasas de incidencia de cáncer en el futuro si se conocen los cambios temporales de la prevalencia de los principales factores de riesgo y pueden predecirse sus posibles cambios en el futuro. Este método se ha utilizado para

Las estadísticas de cáncer actualizadas proporcionan información sobre la magnitud del cáncer para el sistema sanitario en una población. Para poder desarrollar programas a largo plazo de control del cáncer, es necesario predecir cuáles serán las necesidades en el futuro, es decir, es necesario tener estimaciones fiables del número de casos incidentes y prevalentes que se producirán en los próximos años. Los registros de cáncer son una fuente

predecir la incidencia de cáncer de pulmón respecto a la prevalencia del hábito de fumar (como en la Figura 17.3) y el cáncer de mama respecto a los cambios en la fertilidad.

En este caso, como en cualquier otro tipo de predicciones, las interpretaciones deben realizarse con precaución, ya que se basan en un número considerable de suposiciones. Para evaluar su robustez se recomienda proporcionar previsiones en diferentes escenarios posibles como en el ejemplo presentado en la Figura 17.3

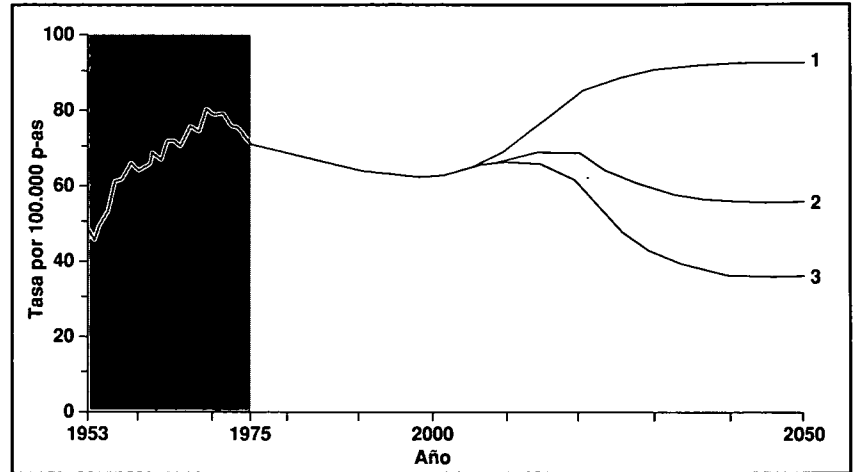


Figura 17.3.

Tasas de incidencia de cáncer de pulmón ajustadas por edad (a la población mundial) en hombres en Finlandia en 1953-1975 y previsiones para las tasas en 1976-2050 obtenidas de un modelo de simulación basado en las siguientes suposiciones: en cada periodo de 5 años consecutivos de 1976-2050, un 10% de cada categoría de fumadores dejará de fumar, y se dará una de estas tres posibles alternativas (tres curvas): (1) un 60% de los no fumadores de 10-14 años, un 30% de los de 15-19 años y un 10% de los de 20-24 años empezarán a fumar; (2) los porcentajes son 30, 15 y 5, respectivamente; y (3) los porcentajes son 15, 7,5 y 2,5, respectivamente (reproducido con la autorización de Oxford University Press, de Hakulinen y Pukkala (1981)).

17.6.2 Evaluación de los programas de control de cáncer *Prevención primaria*

Los registros de cáncer pueden jugar un papel importante en la monitorización y evaluación de la eficacia de las medidas de prevención primaria. Como se ha mencionado en el Apartado 16.2.2, las tendencias en la incidencia de cáncer pueden estar relacionadas con cambios en el tiempo en la exposición a factores de riesgo. Ocasionalmente, cuando la implementación de un programa de prevención primaria se ha limitado a una zona, puede ser posible realizar comparaciones de los cambios en la zona de intervención frente a las zonas «control». Al interpretar estas relaciones hay que tener en cuenta que para que un efecto del cambio en una exposición se vea reflejado en los datos de incidencia del cáncer, es necesario un tiempo considerable (generalmente décadas).

Cribaje y detección precoz

Los registros de cáncer pueden jugar un papel importante en la evaluación y monitorización de programas de cribaje que tienen como objetivo detectar situaciones preinvasivas (p.e. para el cáncer de cuello uterino). Los datos de registros de cáncer se han utilizado en estudios basados en datos rutinarios para examinar tendencias en las tasas de enfermedad respecto a las frecuencias de cribaje dentro de una población, y para comparar las tasas de enfermedad entre diferentes poblaciones en relación a la cobertura ofrecida por sus programas de cribaje (véase Apartado 16.3.1). Por ejemplo, estos estudios han apoyado la hipótesis de que el uso regular de la prueba de frotis de Papanicolaou es eficaz en la reducción de la incidencia de cáncer de cuello uterino.

Los registros de cáncer también pueden contribuir en la identificación de casos en ensayos de intervención y estudios de cohorte diseñados para evaluar programas de cribaje y como fuente no sesgada de casos para estudios caso-control. Los principales aspectos a considerar en el diseño e interpretación de estos estudios se presentaron en el Apartado 16.3.1.

Cuando los programas de cribaje están destinados a la detección de cánceres rápidamente invasivos (p.e. cáncer de mama), la medida principal de su eficacia debe ser la reducción de la mortalidad en lugar de la incidencia. No obstante, una vez se sabe que un programa de cribaje es eficaz, los registros de cáncer pueden ayudar a monitorizar su funcionamiento proporcionando datos sobre las denominadas 'medidas intermedias de eficacia'. La ausencia de un cambio en estas variables intermedias indica que el cribaje no ha sido eficaz. Las estadísticas adecuadas de monitorización proporcionadas por los registros de cáncer son:

- (a) la incidencia de cánceres de intervalo (es decir, cánceres detectados entre tests de cribaje) comparada con la incidencia en la población antes de introducir el cribaje.
- (b) la distribución por estadios de los cánceres detectados por cribaje comparada con la distribución de los cánceres no detectados por cribaje. Si no se observa un desplazamiento de la distribución de estadios hacia las fases iniciales, el programa no es eficaz.
- (c) si el cribaje es eficaz, los cánceres detectados por cribaje deben presentar una mejor supervivencia que los casos no detectados por cribaje.

No obstante, cabe destacar que las variables intermedias están sujetas a numerosos sesgos y por tanto pueden indicar que el programa es eficaz incluso aunque los datos de mortalidad no lo indiquen. Estos aspectos se comentaron en el Apartado 16.3.1.

Prevención terciaria

Las estadísticas de supervivencia se generan a partir del seguimiento (activo o pasivo) de los casos de un registro de cáncer. El análisis de la supervivencia basado en registros poblacionales *no puede* evaluar los tratamientos específicos que sólo pueden realizarse en ensayos clínicos aleatorizados. Por el contrario proporcionan un índice útil para la evaluación del cuidado del cáncer de la zona cubierta ya que estarán incluidos todos los casos de cáncer independientemente del tipo de tratamiento que puedan haber recibido.

Los métodos utilizados en el análisis de la supervivencia son los comentados en el Capítulo 12. El primer requisito para la aplicación de estos métodos es una *definición de caso clara y bien definida*. Ésta debe especificar claramente el tipo de cáncer y/o histología, edad y sexo del paciente y, si se tiene, el estadio de la enfermedad en el momento del diagnóstico. También debe definirse la naturaleza de los casos a incluir – por ejemplo, debe decidirse si se incluyen en el análisis los casos cuya base de diagnóstico más válida sólo es la clínica. También debe decidirse sobre los casos registrados basados sólo en un certificado de defunción (Death Certificate Only, DCO),

para los cuales no se dispone de información sobre la fecha de diagnóstico del cáncer. La práctica más habitual es omitir estos casos del análisis, pero si representan una gran proporción de este tipo de casos, puede ser mejor presentar dos análisis de supervivencia, uno que incluya los casos DCO y otro que los excluya. En cualquier caso, en los informes de supervivencia debe indicarse la proporción de estos casos.

El segundo requisito es un *punto de inicio claro y bien definido*. Para registros de cáncer poblacionales, la fecha de inicio (a partir de la cual se calcula la supervivencia) es la *fecha de incidencia* (véase el Apartado 17.3.1).

El tercer requisito es definir un *punto final del análisis claro y bien definido*. La muerte suele ser el resultado de interés pero algunos registros pueden recopilar datos suficientes para permitir realizar el análisis utilizando como resultado de interés la recidiva del tumor o la primera recurrencia de una complicación concreta. También es necesario formular criterios claros para decidir quién debe considerarse 'perdido del seguimiento'. Por ejemplo, algunos registros supondrán que los casos para los cuales no pueden obtenerse datos de seguimiento durante más de 15 meses deben considerarse 'pérdidas del seguimiento'.

Existen varios problemas en la interpretación de las tendencias temporales en la supervivencia. En primer lugar, las mejorías de la supervivencia pueden ser debidas, como mínimo en parte, a una mejor determinación y archivo de los casos incidentes. En segundo lugar, si se ha observado una tendencia hacia un diagnóstico precoz (p.e. introducción de un programa cribaje), la supervivencia puede mejorar pero la ganancia puede ser debida totalmente a un mayor tiempo de avance en el diagnóstico, sin cambios en las tasas de mortalidad (véase el Apartado 16.3.1).

A pesar de estas limitaciones, las tendencias temporales en la supervivencia son útiles para evaluar en qué grado los nuevos avances terapéuticos han tenido un efecto sobre la población. Por ejemplo, la mejora en la supervivencia observada en ensayos clínicos en el tratamiento de cánceres infantiles parece haberse transferido a la comunidad en muchos países desarrollados, ya que la supervivencia basada en la población para muchos de estos cánceres muestran aumentos significativos en el tiempo (Tabla 17.7).

Cáncer	Supervivencia a 5 años (%)	
	1962-64 ^b	1971-74
Leucemia linfóide	14	39
Enfermedad de Hodgkin	39	79
Linfomas no Hodgkin	20	25
Tumor de Wilms	26	57
Tumores óseos malignos	22	29

^a Datos de Draper *et al.* (1982).
^b 1968-70 para leucemia linfóide.

Tabla 17.7.
Tendencias temporales de supervivencia a 5 años para algunos cánceres infantiles (0-14 años) en Gran Bretaña^a.

Las comparaciones de las estimaciones de supervivencia de cáncer obtenidas a partir de los registros de cáncer poblacionales se utilizan cada vez

más para comparar la eficacia del tratamiento del cáncer en diferentes poblaciones. No obstante, la supervivencia no sólo refleja el tratamiento sino también factores pronósticos tales como estadiaje en el diagnóstico, el tipo histológico y otras características de la enfermedad. Cuando no se dispone de los datos sobre estos factores o cuando su definición no está correctamente estandarizada en los registros, las razones de las variaciones observadas no pueden identificarse correctamente.

Ejemplo 17.4. El registro de cáncer de Khon Kaen en el nordeste de Tailandia es uno de los pocos registros de cáncer poblacionales en países en desarrollo que recopila datos de seguimiento. Estos datos se obtienen a partir de archivos clínicos, certificados de defunción y de un sistema de postales con devolución franqueada enviadas anualmente a cada paciente durante toda su vida. Un total de 10.333 residentes de la provincia de Khon Kaen registrados con cáncer durante los años 1985-1992 se siguieron hasta finales de 1993. La Tabla 17.8 muestra la supervivencia relativa a 5 años para algunos cánceres seleccionados. Estas supervivencias se compararon con los datos de supervivencia estandarizados por edad de dos países desarrollados –los EE.UU. y Escocia (Sriamporn et al., 1995).

Tabla 17.8. Supervivencia relativa a 5 años en la provincia de Khon Kaen (Tailandia), EE.UU. y Escocia^a.

Localización tumoral	CIE-9	Khon Kaen 1985-92	Blancos de EE.UU. 1983-88 ^b	Negros de EE.UU. 1983-88 ^b	Escocia 1983-87 ^b
Estómago	151	23,4	17,2	19,0	12,8
Intestino grueso	153-154	41,9	58,9	50,0	42,3
Hígado	155	9,2	5,9	3,5	4,2
Pulmón	162	15,4	14,7	10,3	7,7
Mama	174	48,1	78,4	61,5	66,8
Cuello uterino	180	60,1	69,2	56,8	61,0
Linfoma no Hodgkin	200, 202	32,5	56,6	49,7	53,2
Leucemia	204-208	19,4	45,7	31,7	41,6

^a Datos de Sriamporn et al. (1995).

^b Estandarizado por la distribución de edad específica del lugar de los sujetos tailandeses.

En el Ejemplo 17.4 la supervivencia en Khon Kaen fue igual o mejor que la de los EE.UU. para los cánceres de estómago, hígado y pulmón. Por tanto, las mejorías en el tratamiento pueden tener poco beneficio en el control de estos cánceres en Tailandia en comparación con los beneficios potenciales que pueden obtenerse de la prevención primaria y secundaria (p.e. control de infección por hepatitis B y contaminación de aflatoxinas para el cáncer de hígado; campañas antitabaco para el cáncer de pulmón). Por el contrario, la supervivencia fue menor para los residentes en Khon Kaen en comparación con los residentes blancos en los EE.UU. para aquellos cánceres cuyo pronóstico está asociado con el diagnóstico precoz (mama, cuello ute-

rino e intestino grueso), indicando que los cribajes pueden proporcionar grandes beneficios potenciales. La supervivencia por leucemia y linfoma también fue menor para los residentes en Khon Kaen, probablemente debido al bajo acceso a los regímenes terapéuticos complejos en estos países.

17.7 Registros de cáncer hospitalarios

Los registros de cáncer hospitalarios son más numerosos y extendidos que los registros de cáncer poblacionales. El primer objetivo de estos registros es contribuir al cuidado del paciente, proporcionando información directa sobre los pacientes con cáncer, el tratamiento que recibieron y sus resultados. También pueden utilizarse para investigación clínica y, en cierto grado, con fines epidemiológicos.

Una de las principales ventajas de los registros hospitalarios es que tienen acceso directo e instantáneo a los archivos médicos, la fuente primaria de los casos. Los tipos de datos recopilados por un registro de hospital también son más amplios que los recopilados por un registro de población. No obstante, existen varias limitaciones de los datos de los registros hospitalarios:

- (1) Están basados en una institución y no en la población general. Esto significa que no se intenta registrar todos los casos de cáncer que aparecen en una población definida; por tanto no pueden determinarse las tasas de incidencia. Los pacientes hospitalizados en más de un hospital se cuentan más de una vez y es posible que la información no esté compartida entre los hospitales que asisten al mismo paciente en diferentes momentos. Los cambios en el tiempo en los números de cualquier tipo de cáncer o de las características del paciente pueden reflejar sólo desplazamientos de pacientes (o médicos) de una institución a otra. Los casos de cáncer en un hospital (o grupo de hospitales) pueden no ser representativos de todos los casos de cáncer que aparecen en una zona. Por ejemplo, algunas instituciones son centros de referencia para tipos específicos de cáncer o para tumores que requieran tratamientos complejos o prolongados.
- (2) La determinación de la muerte es probable que sea menos completa en los registros hospitalarios que en los registros poblacionales debido al acceso y uso limitado de otras fuentes tales como los certificados de defunción y a que raramente se comparte información de este tipo entre hospitales.
- (3) A diferencia de la mayoría de los registros de cáncer poblacionales, los registros hospitalarios no intentan estandarizar los métodos de recogida de datos entre ellos. Por tanto es difícil comparar sus observaciones.

Los registros hospitalarios del cáncer producen informes sobre el número de cánceres observados en el hospital por año y por localización, edad y

sexo. Estos resultados pueden presentarse como *tasas de incidencia proporcional* (es decir, la frecuencia de cánceres de un tipo particular respecto al número total de casos de cáncer – véase el Apartados 4.3.5 y 11.1.6). También pueden proporcionar información sobre métodos de diagnóstico, distribución de estadios, métodos de tratamiento, respuesta al tratamiento y supervivencia en el ámbito institucional.

Los datos de los registros hospitalarios también pueden utilizarse para prever demandas futuras de servicios, equipo y personal en un determinado centro.

Aunque estos registros no pueden proporcionar tasas de incidencia en la población general, pueden utilizarse con fines epidemiológicos. Por ejemplo, pueden realizarse estudios caso-control para investigar la etiología de un cáncer particular por comparación de las características de estos casos con las de un grupo control; este grupo control puede estar formado por pacientes con otros tipos de cáncer o por otros pacientes hospitalizados. El análisis será similar al mostrado en el Ejemplo 17.2.

Cuadro 17.1. Puntos esenciales

- Existen dos tipos principales de registros de cáncer:
 - (a) Los registros de cáncer hospitalarios archivan la información sobre todos los pacientes de cáncer observados en un hospital. Su principal objetivo es monitorizar y planificar el cuidado del paciente a nivel de una institución. No obstante, sus datos poseen un valor limitado en epidemiología debido a que no es posible definir la población de la que proceden los casos.
 - (b) Los registros de cáncer poblacionales recopilan datos de todos los casos nuevos de cáncer que se producen en una población bien definida. Como resultado, y a diferencia de los registros de cáncer hospitalarios, pueden proporcionar datos sobre la aparición de cáncer en una población particular y, por tanto, tienen un valor especial en epidemiología y salud pública.
- Los registros de cáncer poblacionales juegan un papel importante en epidemiología, cuantificando la incidencia y prevalencia de esta enfermedad en la comunidad y como fuente de detección de casos de cáncer en estudios de intervención, cohorte y caso-control. Sus datos también son importantes en la planificación y evaluación de programas de control del cáncer ayudando a establecer prioridades actuales y proporcionando los datos necesarios para prever las necesidades futuras; monitorizando la aparición de cáncer respecto a la prevalencia de factores de riesgo importantes; ayudando a evaluar y monitorizar la eficacia de programas de cribaje y ayudando a evaluar el cuidado del cáncer mediante sus estadísticas de supervivencia.
- Los tipos de datos a obtener en un registro de cáncer poblacional se determinan por sus objetivos, los métodos de recopilación de datos y los recursos disponibles. Debe prestarse especial atención a la calidad de los datos, incluso por encima de la cantidad. La exhaustividad y validez de los datos debe monitorizarse regularmente.
- Los registros de cáncer poblacionales son especialmente útiles en países en desarrollo donde raramente se dispone de datos de mortalidad fiables.

Lecturas adicionales

* Jensen *et al.* (1991) describe con mucho detalle la planificación de registros de cáncer en países desarrollados y subdesarrollados. Describe también los usos de los datos del registro en epidemiología y salud pública.

* Parkin *et al.* (1994) proporciona recomendaciones prácticas sobre cómo los registros de cáncer poblacionales pueden evaluar y monitorizar la calidad de sus datos.

Capítulo 18

Diseño, planificación y realización de la investigación epidemiológica

18.1 Introducción

En capítulos anteriores de este libro, hemos tratado los principios metodológicos básicos del diseño, análisis e interpretación de un estudio. En este capítulo, nos centraremos en algunos aspectos prácticos de como diseñar, planificar y realizar un estudio epidemiológico. Obviamente, sólo podemos comentar aspectos generales. Cada proyecto de investigación y cada situación de estudio presenta problemas únicos que no pueden tratarse aquí.

El primer paso en el diseño de un estudio epidemiológico es tener una 'idea' para un estudio. Éste es un proceso creativo para el cual no pueden darse directivas o consejos. La idea para un estudio normalmente procede del propio trabajo y experiencia y de la consciencia de que existe una necesidad de obtener una respuesta clara a una cuestión de investigación concreta. Una vez se ha identificado la idea para el estudio, el siguiente paso es una revisión crítica de la bibliografía existente para ver qué se sabe exactamente sobre el tema y para estar seguro de que la cuestión no se ha respondido ya. Cuando proceda, deben consultarse los registros de ensayos clínicos y/o directorios de investigación en curso (tales como el de la Agencia Internacional de Investigación sobre el Cáncer). También puede ser útil ponerse en contacto con expertos en el tema. Además de aprender lo que ya se sabe sobre la cuestión, esto ayudará al investigador a familiarizarse con los problemas que han tratado otros investigadores, utilizando diversos métodos de estudio.

Una vez reunida toda la información pertinente sobre el tema, el siguiente paso es plantear la hipótesis del estudio de forma clara y factible. Existe una tendencia a formular hipótesis demasiado amplias y vagas. Por el contrario, deben plantearse en términos de preguntas claras, simples y que puedan responderse. Todas las hipótesis principales y secundarias deben formularse al inicio del estudio para asegurar que se recopilen todos los datos necesarios. Deben incluir una descripción concisa de la(s) exposición(es) o intervención(es) a estudiar, el resultado(s) de interés, y la magnitud del efecto(s) previsto(s) en la población en la que se realizará el estudio.

18.2 Preparación de un protocolo y obtención de financiación

Si estamos convencidos de que vale la pena realizar el estudio y que es factible, el siguiente paso es obtener los fondos necesarios. Si el coste total del proyecto es bajo, puede ser posible realizarlo con recursos ordinarios, pero muchos estudios epidemiológicos son empresas caras que requieren una ayuda económica considerable.

18.2.1 Organismos financiadores

Deben identificarse los organismos financiadores que puedan estar interesados en la investigación propuesta. Algunos organismos financiadores no financian la investigación sobre el cáncer. Otros tienen su propia agenda de investigación y sólo financiarán proyectos que traten las cuestiones concretas en las que están interesados. En algunos países, se publican regularmente directorios de organismos financiadores con sus áreas de interés y presupuesto.

Es importante obtener formularios de solicitud de financiación de todos los organismos financiadores potenciales y leer su material detenidamente, comprobando las áreas en que están especialmente interesados en financiar y los límites de presentación de solicitudes. A veces, su material incluye listas de proyectos financiados en el pasado. Estas listas dan una idea de las áreas que es probable que financien en el futuro y la cantidad de dinero asignado habitualmente a los proyectos individuales. Puede ser beneficioso ponerse en contacto con organismos financiadores en esta fase para comprobar si la propuesta tiene un interés potencial para ellos. Si el estudio propuesto está fuera de sus áreas de interés, no tiene sentido hacer el esfuerzo de preparar y presentar la solicitud.

Redacción de un protocolo

El protocolo del estudio debe estar escrito según las especificaciones del organismo financiador. Aunque la presentación de los formularios de solicitud varía de un organismo financiador a otro, generalmente se dividen en los siguientes apartados:

Hipótesis del estudio

Las hipótesis principal y secundarias del estudio deben redactarse de forma clara y concisa, indicando la(s) exposición(es) y resultado(s) de interés y la magnitud del efecto(s) previsto(s).

Estado de la cuestión y justificación del estudio

Este apartado debe incluir una revisión bastante breve de los conocimientos sobre el tema. Debe «justificar» la necesidad del estudio propuesto indicando claramente su originalidad y el interés potencial de sus resultados. La investigación propuesta puede ser una extensión lógica de trabajos anteriores realizados por los investigadores o de un estudio piloto inicial. En este caso, los resultados de estos estudios deben presentarse en esta sección.

Población a estudiar y métodos de reclutamiento

Deben describirse la situación geográfica y las características demográficas de la población a estudiar. También debe indicarse cualquier razón particular para la elección de la población en estudio. Por ejemplo, una población particular puede haberse elegido para un estudio debido a su exposición excepcionalmente alta a un factor de riesgo concreto o por razones logísticas (p.e. fácil de seguir). Deben presentarse detalles sobre como se reclutarán los sujetos del estudio.

Diseño del estudio

Debe especificarse si será un estudio de intervención, de cohorte, caso-control, transversal o basado en datos de rutina. La elección del diseño tiene que justificarse (véanse Capítulos 5 y 7-11).

Tamaño de la muestra

Es necesario mostrar que el número de sujetos propuesto en el estudio proporcionará el poder o precisión adecuado para detectar o estimar un efecto particular. Las estimaciones del tamaño de la muestra deben presentarse bajo diferentes suposiciones. Una estimación única raramente es convincente (véase Capítulo 15).

Métodos de recogida de datos

Los métodos de recogida de datos (p.e. entrevista, determinaciones de laboratorio, obtención de datos de archivos clínicos, etc.) deben describirse con suficiente detalle para mostrar que el plan se ha pensado en todos sus detalles y que es factible. Deben presentarse las limitaciones prácticas previstas y las estrategias para resolverlas.

Análisis estadístico

Debe presentarse una descripción concisa de los métodos estadísticos a utilizar en el análisis de datos (véanse Capítulos 6 y 14).

Ética

El protocolo de todo estudio que implique a sujetos humanos debe proporcionar respuestas a las siguientes preguntas: ¿Qué se les explicará a los sujetos? ¿Qué ocasionará su colaboración? ¿Se utilizarán procedimientos agresivos? ¿Existen riesgos para los participantes? ¿Cómo se obtendrán los consentimientos? (p.e. a nivel individual o comunitario; consentimiento escrito u oral) ¿Qué pasos se tomarán para asegurar la confidencialidad de los datos? En esta etapa será conveniente consultar las normativas éticas propuestas por el Consejo de Organizaciones Internacionales de Ciencias Médicas (CIOMS) y la Organización Mundial de la Salud (OMS) (1993). La mayoría de los organismos sólo financiarán proyectos que se hayan aprobado por los comités éticos correspondientes y algunos solicitan muestras de la hoja de información y el formulario de consentimiento para los sujetos del estudio.

Calendario

Debe proporcionarse un calendario realista para efectuar las diferentes actividades del estudio. Esto debe incorporar metas a conseguir a intervalos de tiempo regulares (p.e. cada seis meses), que ayudarán a monitorizar el progreso del estudio.

Referencias

En la propuesta de financiación deben incluirse referencias a las publicaciones clave.

Presupuesto

El presupuesto generalmente se divide en los costes de personal, equipo (p.e. ordenadores, congeladores, centrífugas) y gastos corrientes (p.e. gastos de oficina, material informático y de laboratorio, costes de viaje). Cada punto debe estar justificado y debe tener un coste razonable. Esto implicará con frecuencia discusiones con la oficina de personal sobre los costes de personal y la consulta de listas de precios para obtener el precio más barato de suministros y equipo. En algunos países, al calcular el presupuesto final debe tenerse en cuenta la inflación y las fluctuaciones de la moneda.

También es importante comprobar si la institución en la que se basará el estudio añade gastos institucionales. Éstos corresponden a los costes de la institución para administrar la financiación, organizar el pago de salarios, pedir suministros, proporcionar espacio de oficina, electricidad, aire acondicionado, etc. Algunos organismos financiadores se niegan a pagar estos gastos (p.e. Organización Mundial de la Salud), mientras que otros pagarán sólo hasta una cierta proporción del coste total (p.e. la Comisión Europea cubre sólo el 20% del coste total). Si el organismo financiador no paga estos gastos institucionales, algunos de los gastos que normalmente se cubrirían con este apartado, tales como los costes de teléfono, fax y correo, pueden incluirse en el presupuesto.

Difusión de los resultados

Algunos organismos financiadores piden a los solicitantes que indiquen como pretenden informar a los participantes en el estudio sobre los resultados del mismo y cómo se harán llegar a las autoridades sanitarias y comunidad científica pertinentes. Esto se comentará más adelante en el Apartado 18.4.

El Cuadro 18.1 presenta una lista de temas a considerar al preparar solicitudes de financiación. Antes de enviar la solicitud de ayuda al organismo financiador, es útil enviar un borrador a personas con experiencia en escribir y presentar solicitudes para sus comentarios. Aunque nadie puede prever todos los problemas que pudieran surgir, los investigadores experimentados observarán rápidamente muchas dificultades potenciales. En particular, se recomienda incluir a un estadístico en el desarrollo del protocolo del estudio.

El investigador también debe informar a todas las personas cuya autorización o cooperación es necesaria o deseable. La investigación propuesta

en instituciones clínicas o académicas debe presentarse a los jefes de departamento y/o administradores de los hospitales correspondientes. Con frecuencia habrá un comité diseñado especialmente para revisar y aprobar protocolos de estudios. Los estudios realizados en la comunidad deben presentarse a los oficiales sanitarios locales. Además de obtener la autorización solicitada, el investigador puede recibir sugerencias prácticas valiosas y ayuda adicional de estos profesionales, tales como presentaciones a médicos que deberán autorizar el estudio de sus pacientes. El investigador también puede aprender de otras investigaciones similares o relacionadas que se estén llevando a cabo. La cooperación con otros investigadores puede ayudar a evitar la duplicación del esfuerzo y puede permitir compartir los recursos y, posiblemente, incluso los datos.

Cuadro 18.1. Lista de temas a considerar al preparar las solicitudes de financiación

- Identificar los organismos financiadores potencialmente interesados en la investigación propuesta. Obtener los formularios de solicitud de financiación.
- Leer sus instrucciones detenidamente. Comprobar las fechas límite de presentación de solicitudes.
- Comprobar si solicitan la presentación de documentos especiales (p.e. certificado de aceptación a participar en el estudio de todos los colaboradores, confirmación de la aprobación ética) que pueden llevar cierto tiempo conseguirlos.
- Hacer una lista del personal y equipo necesario. Consultar a la oficina de personal para obtener estimaciones de los salarios. Obtener una lista de precios para el equipo. Hacer una estimación de los gastos corrientes (p.e. gastos de oficina, material informático, gastos de viajes). Comprobar si su organización carga gastos institucionales.
- Redactar el protocolo del estudio según las instrucciones del organismo financiador, respetando los diferentes apartados. No exceder el espacio y/o número máximo de palabras especificado para cada uno de los apartados.
- Ser conciso y claro. Comprobar los errores de ortografía detenidamente. Si no escribe la solicitud en su lengua materna puede ser útil enseñárselo a alguien versado en esa lengua.
- Mostrar el borrador del protocolo a expertos en el tema. Revisar el borrador para sus comentarios.
- Asegurarse de enviar la solicitud original y el número solicitado de copias al organismo financiador a tiempo para que llegue a su oficina antes de la fecha de cierre de la presentación de solicitudes.

Revisión de propuestas de financiación

La mayoría de los organismos financiadores envían las solicitudes a expertos en este tema concreto. A estos expertos se les pide evaluar las propuestas respondiendo a las siguientes preguntas: ¿Trata la propuesta una cuestión importante? ¿Es adecuado el diseño del estudio? ¿Están cualificados los solicitantes para realizar la investigación? ¿Es razonable el presupuesto? Cada experto enviará un informe con sus comentarios al organismo financiador.

La mayoría de los organismos financiadores tienen su propio panel de revisión de proyectos de investigación que considerará las solicitudes y los informes de los expertos. El panel clasificará las solicitudes y financiará aquellas en la parte más alta de la escala de clasificación. Ocasionalmente, puede invitarse a los solicitantes a presentar una solicitud revisada que tenga en cuenta los comentarios del experto. La solicitud revisada debe tratar todos los comentarios de los expertos y debe presentarse junto con una carta de introducción detallando las revisiones efectuadas.

18.3 Realización del estudio

Una vez se ha asegurado la financiación del estudio, los investigadores tienen que empezar a planificar la realización del estudio detenidamente y en detalle.

18.3.1 Organización logística

Es importante planificar todas las actividades del estudio desde la fase inicial. La organización logística de un estudio debe regirse por tres principios generales. Primero, deben establecerse etapas realistas al inicio del estudio. Segundo, todos los procedimientos y decisiones del estudio deben documentarse correctamente. Tercero, los gastos deben monitorizarse estrechamente.

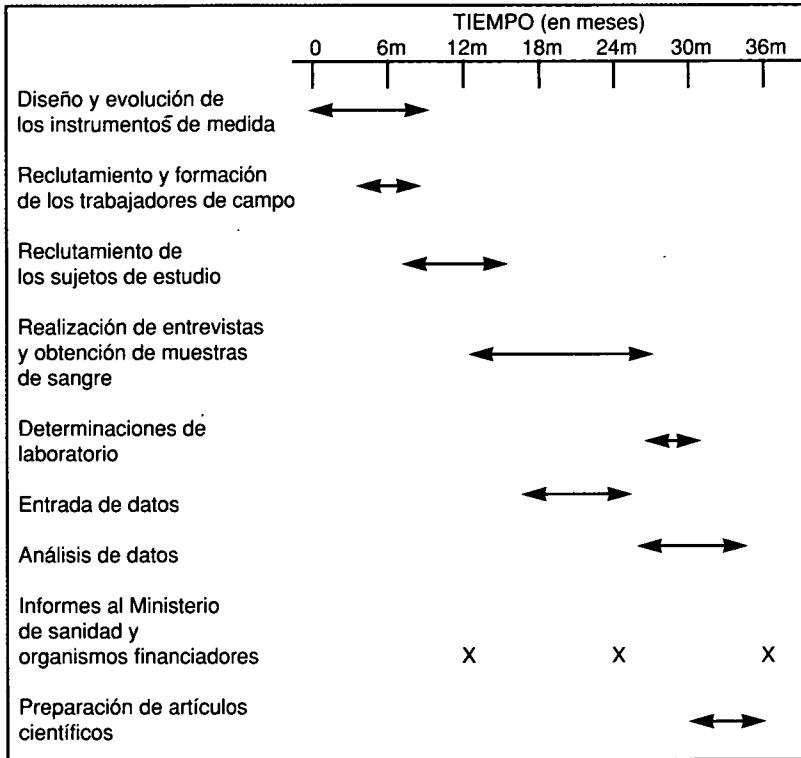
Debe construirse un *calendario* de todas las actividades a realizar. En el Cuadro 18.2 se presenta un ejemplo. El programa debe indicar los objetivos a conseguir a intervalos de tiempo regulares. Debe ser realista. Por ejemplo, debe tener en cuenta los posibles retrasos en el suministro del equipo vital, factores climáticos que puedan afectar al trabajo de campo y las vacaciones y enfermedades del personal.

Debe prepararse un *manual del estudio* en el que se documente con detalle cada procedimiento del estudio (p.e. instrucciones paso a paso para la administración y realización de cuestionarios o para la toma de muestras biológicas). El manual debe actualizarse si se realizan cambios durante el estudio. A todos los miembros del equipo de investigación se les deben dar copias de las partes correspondientes del manual. Además del manual del estudio, también es importante mantener un *diario del estudio* en el que se anoten todos los problemas surgidos y se registren las soluciones adoptadas. Esto será útil para mantener la concordancia de las decisiones a lo largo del estudio.

Debe mantenerse un *registro de gastos* detallado para monitorizar los gas-

tos que se produzcan a lo largo del estudio y facilitar la presentación de los certificados de gastos al organismo financiador. Es importante mantener copias de las nóminas, facturas, cuentas, formularios de pedido, etc. Las cuentas de la financiación pueden monitorizarse fácilmente utilizando una hoja de cálculo informatizada.

Cuadro 18.2. Ejemplo de un calendario de organización



18.3.2 Inclusión, formación y gestión del personal

Los estudios epidemiológicos con frecuencia requieren la inclusión de personal (p.e. estadísticos, entrevistadores, personal de entrada de datos, programadores informáticos, etc.). Para cada persona debe escribirse una descripción del trabajo indicando sus responsabilidades, las cualificaciones mínimas y la experiencia requerida.

Es importante establecer una línea clara de gestión de modo que cada miembro del personal sepa cuáles son sus responsabilidades y a quién debe informar. La correcta formación del personal es fundamental para el éxito del estudio y deben asignarse el tiempo y los recursos adecuados para ello. El funcionamiento de cada miembro del personal debe monitorizarse regularmente y dar opiniones constructivas. Los buenos resultados deben considerarse como un esfuerzo del equipo.

18.3.3 Equipo

El equipo necesario dependerá del diseño del estudio y de la situación en que se efectuará. Los estudios basados en datos de rutina pueden requerir sólo el acceso a unas buenas instalaciones informáticas. Por el contrario, un ensayo de intervención grande realizado en una zona remota puede requerirlo todo, desde equipo informático hasta alojamiento del personal y medios de transporte.

Equipo informático

La elección del equipo informático ('hardware') y los programas informáticos ('software') (véase el Cuadro 18.3) se determinará por el diseño y el tamaño del estudio, por la disponibilidad de consejo técnico para el procesamiento y análisis estadístico, así como por la existencia de empresas locales de servicios. No obstante, la velocidad de desarrollo del equipo informático y de los programas es tal que cualquier consejo o directivas se vuelven rápidamente anticuados. Por tanto, es esencial que se busque el consejo profesional, de un especialista informático en la etapa de planificación del estudio, de modo que se aprovechen los desarrollos más recientes.

Hardware (maquinaria)

Los ordenadores difieren en el tipo de microprocesador, en el tamaño de su memoria de acceso aleatorio (RAM) y en su capacidad de almacenamiento. Los microordenadores más utilizados se basan en los microprocesadores 486 y Pentium. La memoria de acceso aleatorio (RAM) gobierna el tamaño del programa y la cantidad de datos con que la máquina puede trabajar activamente en cualquier momento. Cuanto más grande sea la RAM, mejor. La mayoría de las máquinas tienen como mínimo 4 megabytes (Mb) de memoria de acceso aleatorio (RAM) pero si va a utilizar Windows, es preferible como mínimo 16 Mb. La capacidad de almacenamiento de datos se determina por la capacidad del disco duro. En general, la capacidad del disco duro debe ser como mínimo de 500 Mb, especialmente si van a utilizarse programas de Windows, ya que estos programas son grandes y tienden a utilizar una cantidad considerable de espacio en el disco duro.

Una vez elegido el tipo apropiado de máquina para el estudio, es necesario obtener listas de precios de diferentes fabricantes. Muchos fabricantes diferentes producen máquinas con especificaciones similares y los precios tienden a oscilar considerablemente. Antes de comprar, es importante recordar que el valor real de un objeto es distinto del concepto barato/caro. La fiabilidad del equipo, garantía y servicio de mantenimiento del vendedor son factores importantes que deben considerarse. La disponibilidad de un buen servicio local e instalaciones de reparación es una consideración muy importante, especialmente para estudios en países en desarrollo.

Existen microordenadores con características similares en dos modelos básicos: los ordenadores personales de oficina y los portátiles. Los ordenadores portátiles presentan la ventaja de que pueden trasladarse pero son

más caros que los ordenadores personales de capacidad equivalente. No obstante, debido a que trabajan con baterías recargables evitan la necesidad de un suministro de potencia ininterrumpible [*uninterrupted power supplies (UPS)*] (véase más adelante).

Es absolutamente vital tomar precauciones adecuadas frente a la posibilidad de perder datos que se han introducido en el ordenador. Toda la información del ordenador debe copiarse regularmente en otros dispositivos de almacenamiento para estar prevenido frente a la posibilidad de un fallo del disco duro o del ordenador. Los disquetes son adecuados para almacenar series de datos relativamente pequeñas, pero para series de datos muy grandes pueden ser necesarias las unidades de disco duro extraíbles. Alternativamente, podemos utilizar una cinta para hacer copias de seguridad de los datos en una cinta magnética pequeña. Es esencial hacer copias de seguridad de los datos regularmente, a diario y semanalmente, y crear copias múltiples como copias de seguridad de la copia de seguridad (véase Apartado 18.3.6).

Los suministros de energía no fiables pueden dar lugar a pérdidas de datos y daños graves en el equipo informático. Se recomienda evitar compartir el mismo circuito con otra maquinaria pesada (p.e. motores eléctricos, sistemas de aire acondicionado) ya que hace demandas fuertes pero intermitentes del suministro de potencia. Los estabilizadores de la potencia (o filtros mains) son dispositivos pequeños, a veces incorporados en los enchufes, diseñados para resolver las fluctuaciones rápidas del voltaje. No obstante, no ofrecen protección frente a la pérdida de datos si se produce un fallo de la potencia. Los UPS son dispositivos que proporcionan potencia (de baterías) durante periodos cortos. Por tanto, en el caso de un fallo de la potencia siguen suministrando potencia durante un tiempo hasta que se acaban las baterías. Esto proporciona el tiempo suficiente para grabar el trabajo y desconectar el ordenador sin pérdida de información. Algunos UPS desconectarán el ordenador solos si se prolonga la pérdida del suministro principal de potencia.

El precio de las impresoras varía enormemente. En general, cuanto mejor sea la calidad de impresión, mayor será el precio. Las impresoras matriciales son relativamente baratas y la calidad de su resultado suele ser lo suficientemente buena para la mayoría de las necesidades. Puede ser necesaria una impresora de chorro de tinta o una láser si es necesario producir documentos de alta calidad tales como las solicitudes de financiación, cuestionarios e informes.

El equipo informático puede utilizarse en condiciones ambientales normales. No obstante, se recomienda intentar asegurar que se encuentran en un ambiente exento de polvo con temperaturas inferiores a 25°C (para evitar el sobrecalentamiento) y humedad inferior al 80 por ciento. Por tanto, puede ser necesario un equipo especial para asegurar las condiciones ambientales (p.e. sistemas de aire acondicionado) especialmente en lugares calientes y húmedos.

Cuadro 18.3. Requisitos informáticos.

- Hardware (máquinas)
 - Requisitos mínimos*
 - ordenador
 - impresora
 - Requisitos opcionales*
 - sistemas de copias de seguridad (p.e. cintas) o disco duro extraíble
 - suministro de potencia ininterrumpible (UPS)
- Software (programas informáticos)
 - Requisitos mínimos*
 - sistema operativo
 - programa de base de datos
 - programa estadístico
 - programa procesador de textos
 - programas antivirus
 - Requisitos opcionales*
 - programas de transferencia de datos

Software (programas informáticos)

Para que un ordenador trabaje requiere una serie de instrucciones que controla sus operaciones. Esta serie de instrucciones se denomina sistema operativo. En el pasado, la mayoría de los microordenadores tipo IBM utilizaban el sistema operativo DOS pero hoy el sistema operativo Windows es el más popular. Windows permite trabajar con varios programas simultáneamente (lo que no es posible con el DOS a no ser que se haya instalado un paquete especial de programas) y, por tanto, es más fácil mover datos, textos y figuras entre los diferentes programas.

Aparte del sistema operativo, la mayoría de los estudios epidemiológicos requieren un programa de base de datos para introducir y comprobar los datos, un paquete estadístico para realizar el análisis estadístico, y un procesador de texto para ayudar al diseño y escribir los cuestionarios de recogida y entrada de datos, para producir los informes del progreso del estudio y para escribir los artículos con los resultados finales para su publicación. Existen muchos paquetes informáticos adecuados. El mejor sistema es utilizar paquetes conocidos debido a que éstos ya han sido probados por otros. Con bastante frecuencia los paquetes nuevos contienen «parásitos» (es decir, errores) que pueden causar problemas importantes. Epi Info (versión 6) es un paquete muy útil para el procesamiento de textos, diseño del cuestionario, entrada de datos, validación, tabulaciones simples y análisis estadístico. Se han desarrollado por los Centros de Control de Enfermedades en los Estados Unidos y el Programa Global sobre SIDA en la Organiza-

ción Mundial de la Salud en Ginebra. Este paquete informático se desarrolló específicamente para estudios epidemiológicos, es fácil de utilizar y se distribuye gratis (aparte del envío)^a.

Los ordenadores pueden estar infectados por un «virus», que es una porción de software, escrita por personas sin escrúpulos, que intentan dañar el ordenador o los datos que contiene. Deben tomarse precauciones para evitar que esto ocurra y para detectar inmediatamente cualquier posible contaminación. Se han desarrollado paquetes de programas informáticos de control de virus para comprobar rutinariamente la presencia de virus y limpiarlos del ordenador. Deben escanearse los virus de disquetes antes de transferir datos al ordenador. Es importante asegurar que el programa está actualizado cada vez que aparecen «virus» nuevos.

Si es necesario transferir con regularidad datos entre ordenadores (por ejemplo, de un ordenador portátil a un ordenador de torre), esto puede efectuarse mediante una conexión por cable entre las entradas en serie o en paralelo de las dos máquinas. Será necesario un programa especial para hacer la transferencia de datos. Los datos pueden transferirse mucho más rápidamente de esta forma que utilizando disquetes. Si es necesario un acceso constante a los datos de más de un ordenador, puede crearse una red de ordenadores. No obstante, esto es más complejo de realizar y requiere la intervención de un especialista.

Otro equipo

Puede ser necesario obtener equipo de oficina como mesas, archivos, una fotocopidora, un fax, etc. Los estudios que impliquen la recogida de muestras biológicas pueden requerir equipo de laboratorio como congeladores, centrífugas, etc. (véase el Apartado 18.3.5). Los estudios de campo en zonas remotas pueden requerir el alojamiento del personal y medios de transporte.

18.3.4 Reclutamiento de los sujetos del estudio

El reclutamiento de los sujetos del estudio puede implicar la obtención de listas de sujetos potencialmente elegibles. Estas pueden obtenerse de las autoridades locales (p.e. listas de todos los sujetos residentes en una zona concreta), de médicos u hospitales (p.e. listas de pacientes con una enfermedad concreta), de registros del cáncer, etc. Si es necesaria una lista de todos los residentes de una zona geográfica particular como base de muestreo y no se dispone de listas de recuento o las disponibles son muy inexactas u obsoletas, será necesario realizar un censo de la población definida en estudio antes de reclutar los sujetos del estudio.

El reclutamiento de los sujetos normalmente implica escribir cartas y/o encontrarse con los sujetos del estudio, sus médicos, representantes sanitarios locales, jefes de las comunidades, etc. para explicar los objetivos del estudio y para obtener su apoyo y consentimiento.

Una vez reclutado en el estudio, a cada sujeto se le debe asignar un número único de identificación del estudio. Este número debe permanecer

^a: Más detalles pueden obtenerse de USD Incorporated (2075A West Park Place, Stone Mountain, GA 30087, USA, Tel. +1 404 469 4098).

constante durante el estudio. Ayudará a mantener la confidencialidad y el ciego y se utilizará para relacionar la información del individuo obtenida de fuentes diferentes (p.e. cuestionarios, archivos clínicos y formularios de laboratorio).

18.3.5 Obtención de muestras biológicas y determinaciones de laboratorio

Muchos estudios epidemiológicos relacionados con el cáncer implican la recogida de muestras biológicas tales como muestras de sangre, suero, plasma, orina, células y tejido de los participantes en el estudio.

Obtención y procesamiento de muestras biológicas

Las muestras biológicas pueden tomarse específicamente para el estudio epidemiológico o pueden proceder de muestras tomadas con fines clínicos. Las actividades clínicas pueden ser una fuente fácil de materiales biológicos pero las muestras pueden no ser adecuadas para una investigación epidemiológica, especialmente si las han recogido diferentes personas con métodos de recogida y procesamiento diferentes. La cantidad de muestra disponible puede no ser suficiente en volumen o calidad, las muestras pueden haber permanecido demasiado tiempo en condiciones de conservación subóptimas o pueden haberse contaminado con materiales extraños (véase Apartado 18.4).

La sangre puede obtenerse por punción venosa o por pinchazo en el dedo (o en el talón). Los pinchazos en el dedo (o en el talón) tienden a ser más aceptables para los participantes en el estudio. El volumen de sangre obtenido con este método, aunque menor que el obtenido por punción venosa, suele ser suficiente para la mayoría de las pruebas de laboratorio. Por tanto, es importante comprobar con el personal del laboratorio la cantidad mínima de sangre que es necesario extraer para realizar todas las pruebas de laboratorio necesarias. El horario de la extracción de sangre también puede ser importante (p.e. tiempo desde la exposición, hora del día, momento dentro del ciclo menstrual, etc.). Algunos ensayos de laboratorio requieren muestras en ayunas. Otros factores tales como la postura del sujeto y el uso de un torniquete también pueden afectar la concentración de muchos componentes de la sangre.

Tras la recogida, la sangre puede separarse en varios componentes incluyendo el suero, plasma, eritrocitos y leucocitos. La separación debe realizarse poco después de la extracción de la sangre (idealmente, dentro de las 2 horas siguientes). Durante este periodo, las muestras pueden dejarse a temperatura ambiente, pero preferentemente a 4°C. Las muestras de suero y plasma deben congelarse lo antes posible después de la separación y conservarse a -20°C. Esto es adecuado para la mayoría de ensayos, como mínimo durante varias semanas. Algunas pruebas requieren la conservación inmediata a menos de -70°C (p.e. determinaciones de vitamina C).

Algunos ensayos de orina requieren la recogida de muestras de orina durante 24 horas mientras que otros pueden realizarse en una muestra única

tomada 'en el momento'. La recogida de muestras de orina durante 24 horas es logísticamente más compleja y más difícil de aceptar por los sujetos del estudio. Es necesario prestar especial atención e instruir a los sujetos del estudio en la toma de muestras de orina durante 24 horas y al comprobar que se han seguido las instrucciones. Normalmente es necesario dejar el recipiente con el individuo toda la noche y acordar recogerlo el día siguiente. Deben medirse los niveles de creatinina para tener en cuenta y corregir las determinaciones de posibles pérdidas de orina. Una buena comprobación de la totalidad de la recogida de orina es la administración de 250 mg de ácido para-aminobenzoico, que se recupera completamente en la orina en las 24 horas siguientes a su administración.

Antes de almacenar las muestras durante periodos de tiempo prolongados puede ser necesario añadir fijadores y estabilizantes apropiados. Esto debe comprobarse con el personal del laboratorio. Las muestras biológicas se alteran fácilmente por la congelación y descongelación repetidas. Por tanto, las muestras deben dividirse en pequeñas porciones antes de congelarlas. El tamaño ideal de las alícuotas debe ser tal que cada una contenga el material suficiente y justo para realizar las pruebas necesarias en un momento concreto.

Todas las muestras biológicas deben etiquetarse correctamente. Para asegurar la confidencialidad, la etiqueta debe contener el número único de identificación del estudio del sujeto pero no otros detalles personales (p.e. nombre). El número de identificación permitirá relacionar los resultados de las pruebas del laboratorio con los archivos del individual del cual se tomaron las muestras. En algunos casos, será apropiado incluir en la etiqueta la fecha de recogida y el tipo de muestra, si no es evidente. No obstante, es importante asegurar que el personal del laboratorio se mantenga ignorante respecto al estado de exposición (o resultado) de los sujetos de los que se tomaron las muestras. Existen en el mercado etiquetas adhesivas preimpresas con cada número de identificación por duplicado varias veces. También pueden hacerse fácilmente con un ordenador. Deben utilizarse bolígrafos resistentes al agua excepto si las muestras se van a congelar en mezclas salino-alcohólicas, en cuyo caso deben utilizarse lápices normales. El etiquetado debe realizarse en el recipiente y nunca sólo en el tapón. Los números y letras deben escribirse de forma clara y estandarizada (véase el Cuadro 18.5). Es importante probar las etiquetas para ver como se adhieren y como se mantiene el escrito en bolígrafo/lápiz durante el transporte y conservación.

Las muestras pueden conservarse temporalmente en un laboratorio situado donde se está efectuando el trabajo de campo antes de enviarlas al laboratorio donde se van a realizar las pruebas y/o a su lugar de almacenamiento permanente. Si se toma y almacena una gran cantidad de muestras es importante conservarlas de forma que permita la recuperación rápida de cualquier muestra en concreto. Por ejemplo, las muestras pueden conservarse en lotes según la fecha en que se tomaron o congelaron. Esta información puede computerizarse de modo que cualquier muestra en particular pueda localizarse fácil y rápidamente.

Cuadro 18.4. Recogida, procesamiento, transporte y almacenamiento de muestras biológicas

- En el manual del estudio deben darse instrucciones paso a paso sobre como tomar, procesar, transportar y conservar las muestras biológicas.
- Todos los métodos para la recogida, procesamiento, conservación y transporte de muestras deben probarse a escala piloto antes de implementarlos.
- Todas las muestras biológicas deben considerarse potencialmente infecciosas. Todos los implicados en su recogida, procesamiento, análisis o conservación deben tomar las medidas necesarias para protegerse frente a la posibilidad de infectarse (p.e. llevando guantes desechables).
- La calidad de las determinaciones de laboratorio debe monitorizarse estrechamente. Esto consiste en:

Control de calidad interno

- Utilizar controles estándar internos.
- Realizar determinaciones por duplicado.
- Enviar una muestra aleatoria de las muestras dos veces, sin que el personal del laboratorio sepa que son duplicados.

Control de calidad externo

- Enviar muestras aleatorias por duplicado a un laboratorio de referencia.

Control de calidad de los procedimientos del laboratorio

Los laboratorios deben mantener un registro de los reactivos, equipos de ensayo, material de laboratorio, los lotes de suministros y reactivos utilizados en diferentes momentos y el número de veces que se ha descongelado y vuelto a congelar cada alícuota. También debe realizarse un registro de las razones por las que una muestra se considera 'inaceptable' (p.e. material insuficiente, procesamiento inadecuado, conservación o transporte) y de acontecimientos poco habituales que pueden afectar a los resultados de un ensayo (p.e. fallos de la potencia, errores en los procedimientos de ensayo).

Muchos análisis de laboratorio realizan 'controles internos' a partir de reactivos estandarizados de concentración o cantidad conocidas. Además, muchos análisis se realizan por duplicado como procedimiento de rutina normal del laboratorio. Aunque los controles internos y las determinaciones por duplicado son importantes para evaluar la calidad de las determinaciones de laboratorio, no son suficientes. Debe enviarse al laboratorio una muestra aleatoria de las muestras y analizarlas dos veces, sin que el personal del laboratorio sepa cuáles son las muestras duplicadas. La reproducibilidad de resultados entre la primera y la segunda determinaciones proporciona una indicación de la calidad interna de los procedimientos del laboratorio (véase el Apartado 2.6.2). Idealmente, la reproducibilidad debe

comprobarse dentro de lotes, entre lotes y de un periodo de tiempo a otro (p.e. cada semana). Es importante medir la reproducibilidad intra-observador (el mismo analista procesa muestras por duplicado varias veces), inter-observador (dos analistas diferentes procesan las mismas muestras independientemente) e inter-producto (análisis de las mismas muestras con diferentes lotes de reactivos). No obstante, una reproducibilidad alta no asegura una validez alta de las determinaciones. Por tanto, es recomendable enviar una muestra aleatoria de duplicados de muestras a un laboratorio de referencia externo, cuyas determinaciones se tomarán como 'patrón de oro' (véase el Apartado 2.6.1).

Los resultados de laboratorio deben registrarse en los libros del laboratorio o, preferentemente, en cuestionarios preparados especialmente para facilitar la entrada al ordenador. Es importante asegurar que los cuestionarios estén diseñados de forma que permitan el registro de problemas o aspectos concretos que pueden ser importantes para la interpretación de la determinación. Por ejemplo, debe incluir aspectos sobre la identificación de los técnicos implicados en cada ensayo (para comprobar la variación inter-observador), el lote de los reactivos utilizados (para comprobar la variación inter-producto) y sobre problemas técnicos que puedan surgir (p.e. muestras perdidas y rotas, muestras con material insuficiente, error en los procedimientos de ensayo, etc.).

18.3.6 Procesamiento y edición de datos

Los diferentes pasos del procesamiento y edición de datos deben planificarse en la fase de diseño del estudio. Esto debe realizarse consultando con un estadístico y un programador informático.

Diseño de cuestionarios de recopilación de datos

Todos los cuestionarios a utilizar para el registro de datos en un estudio deben diseñarse y probar previamente para asegurar que se obtienen los datos apropiados y que pueden extraerse fácilmente para su procesamiento. Este principio general es aplicable a todos los cuestionarios de recogida de datos: cuestionarios, formularios de laboratorio, formularios de extracción de datos de notas de hospitales, etc. Estos aspectos se comentaron en el Capítulo 2 y el Anexo 2.1.

Recopilación de datos y archivo

La formación adecuada es el aspecto más importante del control de calidad de los datos. Debe centrarse en la exactitud y exhaustividad, con especial atención en mostrar cómo y por qué el trabajo de todo el proyecto depende de la calidad de los datos archivados. Debe enseñarse a los trabajadores de campo cómo escribir los números y letras de modo que el personal de codificación y de entrada de datos no tenga dificultades en la lectura de los formularios. Algunas letras y números se confunden con frecuencia y deben darse instrucciones especiales respecto a éstos. En el Cuadro 18.5 se muestran ejemplos.

Cuadro 18.5. Números y letras escritos a mano que suelen confundirse y soluciones posibles para evitar la confusión

Caracteres confundidos	Solución
1 y 7	Esta confusión aparece si los unos se escriben con un trazo inicial en la parte superior (por ejemplo 1). Si es éste el caso, se deben escribir siempre los setes con una línea horizontal a través (como hacen los franceses), es decir 7̄. Una solución más simple puede ser insistir en que los unos se escriban con un trazo único (por ejemplo /).
O y 0 (cero)	Nota 1. Escribir 0 (cero) con una línea que lo cruce, es decir 0̄
4 y 9	Escribir 4 como 4̄
4 y 7	¡Estos dígitos se pueden confundir si los siete se escriben con una línea horizontal! Instruir a los trabajadores para asegurar que la parte superior del siete se escribe horizontal.
6 y 9	Importante cuando se codifican muestras de laboratorio (por ejemplo, ¿es 61 ó 19?). Trazar una línea horizontal debajo de todos los números 19̄.
2 y Z	Nota 1. Escribir siempre la Z con una línea horizontal a través, es decir Z̄
5 y S	Nota 1. Escribir siempre el 5 con dos trazos de bolígrafo.
O y Q	Nota 1
l y 1	Nota 1. Escribir siempre l con 'sombbrero y zapatos', no en un trazo único, es decir l̄ y no l
U y V	Evitar las dos letras como códigos siempre que sea posible (pero serán necesarias para nombres).

Nota 1. Evite utilizar el carácter alfabético en campos de datos que puedan contener información alfabética o numérica.

^a Reproducido con la autorización de Smith & Morrow, 1996)

Debe evitarse dejar blancos ya que un blanco puede interpretarse de diferentes formas. Un blanco puede significar que la pregunta no procedía, no se sabía la respuesta o, simplemente, que el entrevistador se había saltado la pregunta por error. Además, algunos programas de ordenador interpretan los blancos como ceros, que puede ser un código válido. Es preferi-

ble diseñar los cuestionarios de modo que se prevean todas las opciones (por ejemplo, las respuestas a una pregunta específica pueden ser: 'sí', 'no', 'no se sabe' y 'no procede').

La información debe anotarse a lápiz o con un bolígrafo. Los errores deben corregirse escribiendo la respuesta correcta por encima o debajo de la respuesta original pero nunca escribiendo encima de la respuesta incorrecta. Los trabajadores de campo y supervisores pueden utilizar bolígrafos de colores diferentes de modo que esté claro donde se han hecho correcciones.

Las actividades de recogida de datos deben supervisarse y la calidad de los datos monitorizarse regularmente. Cada actividad de recogida de datos debe diseñarse de tal forma que pueda comprobarse. Las comprobaciones deben estar diseñadas con el objetivo de detectar errores en lugar de probar la alta calidad de los datos.

Debe comprobarse la exhaustividad de los datos y la consistencia interna de los cuestionarios y demás (Cuadro 18.6). El trabajo de los entrevistadores debe monitorizarse regularmente y deben darse consejos constructivos. Si es posible, el mismo entrevistador, colaborador y/o supervisor deberían volver a entrevistar a una muestra aleatoria de sujetos para evaluar la reproducibilidad de las respuestas.

Regularmente deben hacerse tablas intermedias y gráficas de dispersión de las variables más importantes. Estas pueden ayudar a detectar problemas. Por ejemplo, patrones sistemáticos en los datos (p.e. un entrevistador con una tasa mucho mayor de rechazos que los otros) indicarán la necesidad de comprobar aspectos concretos de la recogida de datos.

Cuadro 18.6. Asegurando la buena calidad de los datos

- Asegurar la correcta formación y supervisión de los trabajadores de campo.
- Comprobar muestras de formularios de recogida de datos para evaluar su exhaustividad y exactitud.
- Evaluar al entrevistador observando/escuchando la entrevista.
- Volver a entrevistar una muestra aleatoria de sujetos. La segunda entrevista la puede realizar el supervisor, otro entrevistador, o el mismo entrevistador. Evaluar la reproducibilidad de los datos obtenidos en las dos entrevistas.
- Tabular las variables más importantes por entrevistador, para evaluar la variabilidad entre entrevistadores.

Codificación

La codificación es el nombre dado al procedimiento utilizado para convertir los datos de un cuestionario de recogida de datos en un formato ade-

cuado para el análisis en ordenador asignando un código numérico a cada respuesta posible en el cuestionario de recogida de datos. Por ejemplo, el sexo puede codificarse 1 para hombres y 2 para mujeres; sólo se introducirán el número 1 ó 2 en el archivo del ordenador. Los datos numéricos (p.e. número de hijos) no requieren la codificación ya que puede introducirse el número exacto.

Muchos cuestionarios de recogida de datos están precodificados, es decir, están diseñados de forma que a cada posible respuesta se le asigna un código en el cuestionario (ver apéndice 2.1). El trabajador de campo elige una (o más) de estas opciones precodificadas. Estos cuestionarios precodificados presentan la ventaja de estar casi listos para la introducción en el ordenador por el personal de entrada de datos y de reducir los errores de transcripción. No obstante, todavía es necesaria la codificación posterior para preguntas abiertas y para respuestas que no se ajustan a las categorías precodificadas. A estas respuestas se les pueden asignar códigos adicionales si es necesario.

Una estrategia alternativa es codificar los datos sólo después de que se haya terminado el cuestionario. Los cuestionarios de recogida de datos están diseñados de tal forma que existe una columna específica para la codificación en la derecha de la página. A los trabajadores de campo (o los sujetos del estudio si el cuestionario es auto-administrado) se les pide que ignoren esta columna y rellenen las respuestas en los espacios proporcionados a la izquierda. Los trabajadores de campo o el personal de codificación especialmente formado codifican más tarde las respuestas y los códigos se escriben en la columna derecha. El tiempo entre la recogida de datos y la codificación debe ser el menor posible de modo que puedan detectarse incongruencias en los datos y pedir al trabajador de campo que se ponga en contacto con el sujeto para corregirlas.

Algunos aspectos requieren especial atención al desarrollar las instrucciones de codificación. En primer lugar, las variables numéricas tales como el número de hijos, peso, altura no deben codificarse en categorías predefinidas. Deben registrarse sus valores reales para mantener todos los detalles de los datos en el análisis. Los datos pueden agruparse a continuación de diferentes formas, si procede. En segundo lugar, deben darse códigos específicos si una pregunta concreta no procede a un entrevistado (p.e. número de embarazos a un hombre) o no se sabe la respuesta o el trabajador de campo dejó en blanco la respuesta por error (por ejemplo, 1 = sí; 2 = no; 7 = no registrado; 8 = no procede; 9 = no se sabe). Asimismo, deben desarrollarse códigos de resultado específicos en los formularios de laboratorio para permitir la codificación de problemas técnicos tales como muestras rotas, material insuficiente, error en los procedimientos de laboratorio, etc. Para reducir los errores de codificación se recomienda, si es posible, que el significado de un código concreto permanezca constante en el cuestionario de recogida de datos (p.e., debe asignarse el código 9 a la respuesta de 'no se sabe' de todas las preguntas). En tercer lugar, debe prepararse un manual de codificación con todos los valores y normas de codificación y distribuirlo a

los trabajadores de campo y al personal de codificación y de entrada de datos. Este manual debe actualizarse si se producen cambios en la codificación durante el estudio.

Introducción de datos en el ordenador

Una vez codificados los datos, la información de los cuestionarios está lista para introducirse en un ordenador. Para cada tipo de cuestionario de recogida de datos, debe crearse un fichero de ordenador para introducir los datos. Puede realizarse un programa de entrada de datos de forma que la pantalla del ordenador se parezca a la presentación del cuestionario de recogida de datos para reducir errores del personal de entrada de datos. El programa de entrada de datos suele desarrollarse con un paquete de base de datos. La mayoría de estos programas incorpora controles de concordancia y de rangos de forma que los datos se comprueban y editan de forma interactiva. Por ejemplo, si se introduce un valor inaceptable aparece un mensaje de error en la pantalla y una señal audible avisa del error a la persona que introduce los datos. También es importante diseñar los programas de entrada de datos de forma que no permita la entrada de datos de registros a los que erróneamente se les asignó el mismo número de identificación.

A cada variable para la cual hay datos registrados se le debe asignar un nombre para su uso en el procesamiento por ordenador. Las variables pueden identificarse por números (p.e. 'VAR001') pero es mejor usar nombres abreviados que identifican fácilmente a las variables, tales como 'sexo', 'edad', 'hijos', 'eprimnac' (para edad en el primer nacimiento). El número de caracteres a utilizar en el nombre de una variable no debe exceder el número permitido por los programas (la mayoría de los paquetes no permite el uso de más de 8 caracteres).

Si los datos sobre el mismo individuo pueden haberse obtenido de diferentes fuentes (p.e. cuestionario administrado por el entrevistador, notas clínicas e informes de laboratorio) o estudios de seguimiento realizados a diferentes momentos en el tiempo, es mejor grabar la información de cada fuente o estudio en un fichero separado. Es relativamente fácil relacionar datos de diferentes ficheros de bases de datos utilizando un código común de identificación de la persona (es decir, el número único del estudio).

Cada serie de datos recopilada en un estudio debe introducirse dos veces en el ordenador. La segunda entrada debe ser independiente de la primera y la debe realizar otra persona. Los dos ficheros de datos deben compararse mediante un programa de ordenador diseñado especialmente (disponible en Epi Info). Cualquier diferencia detectada por este programa debe comprobarse frente a los registros originales y editar los ficheros incorrectos. Debe volver a pasarse el programa para comparar los ficheros y repetir el proceso de edición hasta que no se encuentren diferencias entre los ficheros.

Debe prepararse un manual con los nombres de los ficheros en los que se han introducido los datos de cada cuestionario y con detalles de todas las variables (qué es la variable, significado de los códigos, valores permitidos,

etc.) y distribuirlo a todo el personal de procesamiento de datos. Si es necesario debe actualizarse regularmente.

Limpieza de datos

Para cada variable deben realizarse comprobaciones del rango de valores y de concordancia, ya sea cuando se introducen en el ordenador o cuando ya se han introducido. Las comprobaciones de intervalo identifican valores inaceptables. Por ejemplo, si la variable sexo estaba codificada '1' para hombres y '2' para mujeres, los valores diferentes de 1 y 2 se marcarán como errores. Los controles de concordancia hacen comprobaciones cruzadas de la concordancia de datos para variables relacionadas. Por ejemplo, los hombres no pueden tener un diagnóstico de cáncer de ovario.

Además de comprobar los valores incorrectos o no habituales, debe examinarse la distribución de cada variable. Las tablas intermedias de los datos y gráficas de dispersión de variables cuantitativas, tales como altura y peso, son eficaces para mostrar observaciones poco probables. Éstas deben compararse con los cuestionarios de recogida de datos ya que los valores son poco probables pero no son necesariamente imposibles. En algunos casos, puede ser posible corregir los datos. En otros, puede ser necesario insertar un código de «valor ausente» si es seguro que los datos son erróneos (p.e. una altura imposible).

Actualización y almacenamiento

Deben hacerse regularmente copias de seguridad de los datos, lo ideal es a diario, pero si no es posible, como mínimo cada semana. Como mínimo deben existir siempre dos copias de seguridad de todos los datos del ordenador. Algunas de las copias deben guardarse en un lugar geográficamente separado, en un ambiente seco y relativamente exento de polvo para prevenir acontecimientos como fuego, inundaciones, robo, etc. Debe realizarse una lista completa y actualizada de los datos de los que se mantienen copias en todos los discos y cintas almacenados y en más de un lugar.

Los cuestionarios de recogida de datos también deben almacenarse en un lugar seguro para garantizar la confidencialidad de los datos. Los cuestionarios deben archivar en lotes o en orden seriado (por fecha de recogida o por número de identificación del estudio) para facilitar su recuperación para comprobar errores aparentes detectados en los archivos del ordenador. Lo ideal sería que los cuestionarios se mantuvieran durante un largo período después del final del estudio pero las limitaciones del espacio pueden impedirlo. En estudios grandes, puede considerarse la posibilidad de copiar los cuestionarios de recogida de datos en microfichas.

Preparación de datos para el análisis

Las variables introducidas en los ficheros del ordenador no siempre son las adecuadas para el análisis de los datos. Puede ser necesario volver a codificar y computar variables nuevas. Por ejemplo, calcular el índice de masa corporal (IMC) a partir de la altura y peso o la edad en el momento del es-

tudio a partir de la fecha de nacimiento y la fecha de la entrevista. Además, las variables continuas normalmente se agrupan cuando se utilizan los métodos «clásicos» de análisis basados en la estratificación. También puede ser necesaria la combinación de grupos para variables categóricas con muchas categorías o para categorías en las que no hay casos, o muy pocos.

Es preferible mantener una copia del fichero original (tal y como se introdujo a partir de los cuestionarios de recogida de datos) y crear ficheros nuevos para las versiones nuevas de los datos. Los nombres de versiones consecutivas del mismo fichero deben elegirse de forma que sea fácil identificar cual contiene la última versión (p.e., STUD1, STUD2, STUD3, etc.). Es importante mantener copias de los programas escritos para recodificar variables y computar variables nuevas como documentación de lo que se realizó exactamente y para permitir crear nuevas versiones de los datos a partir del fichero original si se detectan errores.

Debe prestarse especial atención a la planificación del análisis de los datos (véanse los Capítulos 3, 4, 6 y 14). Los pasos indicados en el Cuadro 14.1 deben utilizarse como directiva general.

18.4 Difusión de los resultados

Los proyectos de investigación sólo valen la pena si sus resultados se difunden correctamente en la comunidad científica. Esto normalmente implica la publicación de los resultados en revistas científicas acreditadas. Los resultados también pueden presentarse en conferencias y encuentros científicos. Dependiendo de los objetivos y la importancia del estudio, puede ser adecuado producir informes más detallados y/o informes escritos en un lenguaje simple para distribuirlo a las autoridades sanitarias, jefes de comunidades, sujetos del estudio, etc.

18.4.1 Redacción de un artículo

El proceso de redacción un artículo científico tiene muchas similitudes con el de escribir una solicitud de financiación. Así pues, la mayoría de los temas presentados en el Cuadro 18.1 también son aplicables para la preparación de artículos científicos.

Debe empezarse identificando las revistas que puedan estar interesadas en el tema de nuestro estudio y sean adecuadas a la importancia de los resultados. El Cuadro 18.7 muestra algunas de las revistas más importantes para los epidemiólogos del cáncer. Una forma útil de elegir una revista adecuada es ver dónde se ha publicado la mayor parte del trabajo sobre el mismo tema.

Deben obtenerse los detalles de los requisitos de la revista, estilo, títulos, sistema de referenciar, etc., y seguirlos estrictamente. En la mayoría de los ejemplares de las revistas se suelen publicar las 'instrucciones para los autores'. El borrador del artículo debe prepararse exactamente según las instrucciones. Es importante pensar detenidamente sobre cómo ilustrar los resultados, tablas, figuras, etc., de forma que los principales resultados del estudio se presenten claramente a los lectores potenciales.

La mayoría de las revistas pedirán a los autores que estructuren el artículo en los siguientes apartados:

Página título

Esta página debe incluir un título informativo y conciso para el artículo; el nombre(s) del autor(es); el nombre de la institución donde se ha realizado el trabajo; la filiación institucional de cada autor; y el nombre y dirección del autor al que debe dirigirse la correspondencia.

Resumen

El resumen debe incluir en no más unos cientos de palabras la razón del estudio, los principales resultados y las conclusiones principales. Algunas revistas (p.e. *British Medical Journal*) piden un resumen estructurado que se divide en apartados especiales. Es útil obtener algunos ejemplos de resúmenes publicados en la revista antes de ponerse a escribir uno. Algunas revistas pedirán a los autores que proporcionen palabras clave de los *Medical Subject Headings* del *Index Medicus* para indexarlo.

Introducción

Este apartado debe incluir una descripción de la historia de la investigación (citando sólo referencias esenciales) y las razones para realizar el estudio.

Datos y métodos

Los métodos (incluyendo el método estadístico) utilizados deben describirse con suficiente detalle de forma que alguien que lea el artículo pueda repetir el estudio. En la investigación con sujetos humanos, es necesario destacar las implicaciones éticas que puedan surgir de la realización del estudio, si se obtuvo una aprobación de los comités éticos correspondientes y cómo se obtuvo el consentimiento de los participantes en el estudio.

Resultados

Los resultados deben presentarse de forma clara y concisa. Deben evitarse las repeticiones presentando los datos en un solo formato, es decir, los resultados representados en una tabla no deben presentarse en el texto.

Discusión

La discusión debe cubrir los aspectos presentados en el Capítulo 13: aspectos relevantes y limitaciones del diseño del estudio, calidad de los datos, sesgo, confusión y azar. Por último, los resultados del estudio deben comentarse en el contexto de trabajos anteriores.

Agradecimientos

Deben agradecerse las contribuciones, incluyendo ayuda técnica u otra ayuda, y los organismos financiadores que han proporcionado el apoyo económico para el estudio.

Referencias

Las referencias deben escribirse según las especificaciones de la revista en la que va a presentarse el artículo.

Tablas y figuras

En el Apartado 3.4 se presentaron las normas generales para la preparación de tablas y gráficos. Las revistas pueden tener normas específicas, que deben seguirse detenidamente.

Antes de presentar el artículo, es importante enviar un borrador a colegas que suelen publicar artículos y pedirles sus comentarios. Si es necesario, se debe revisar el manuscrito antes de enviarlo a la revista, para que hagan sus comentarios.

18.4.2 Revisión

Si el editor considera que la revista puede aceptar el artículo, lo enviará a revisores expertos. Los informes de estos revisores determinarán si el artículo se acepta o rechaza. Los artículos raramente se aceptan sin modificaciones. Normalmente se rechazará o se aceptará una vez modificado de acuerdo a los comentarios de los revisores. Los autores deben presentar una nueva versión del artículo que los tiene en cuenta junto con una carta de introducción al editor explicando exactamente los cambios introducidos

Lecturas adicionales

* El libro de Smith y Morrow (1996) proporciona información práctica útil sobre cómo planificar y realizar estudios epidemiológicos. Aunque el libro se centra en estudios de campo, la mayoría de los temas presentados son aplicables a cualquier tipo de estudio epidemiológico.

Cuadro 18.7. Lista de algunas revistas internacionales relevantes para los epidemiólogos de cáncer

<i>American Journal of Epidemiology</i>	<i>International Journal of Cancer</i>
<i>British Journal of Cancer</i>	<i>International Journal of Epidemiology</i>
<i>British Medical Journal</i>	<i>Journal of Epidemiology and Community Health</i>
<i>Cancer</i>	<i>Journal of National Cancer Institute</i>
<i>Cancer Causes and Control</i>	<i>Lancet</i>
<i>Cancer Research</i>	<i>Statistics in Medicine</i>
<i>Epidemiology</i>	
<i>European Journal of Cancer</i>	

Referencias

- Acheson, E.D. (1967) *Medical Record Linkage*, Oxford, Oxford University Press
- Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New Engl. J. Med.*, **330**, 1029-1035
- Altman, D.G. (1991) *Practical Statistics for Medical Research*, London, Chapman & Hall
- American Cancer Society (1951) *Manual of Tumor Nomenclature and Coding*, first edition. Washington, DC
- Andersen, L.P., Rosenstock, S.J., Bonnevie, O. & Jørgensen, T. (1996) Seroprevalence of immunoglobulin G, M, and A antibodies to *Helicobacter pylori* in an unselected Danish population. *Am. J. Epidemiol.*, **143**, 1157-1164
- Anon. (1901) Verhandlungen des Comités für Krebsforschung. *Dtsch. Med. Wochenschr.*, suppl., 305-312
- Aristizabal, N., Cuello, C., Correa, P., Collazos, T. & Haenszel, W. (1984) The impact of vaginal cytology on cervical cancer risks in Cali, Colombia. *Int. J. Cancer*, **34**, 5-9
- Armitage, P. & Berry, G. (1994) *Statistical Methods in Medical Research*, third edition, Oxford, Blackwell Scientific Publications
- Armstrong, B. & Mann, J.I. (1985) Diet. In: Vessey, M.P. & Gray, M. (eds), *Cancer Risks and Prevention*, Oxford, Oxford University Press, p. 71
- Armstrong, B.K., White, E. & Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology* (Monographs in Epidemiology and Biostatistics, Vol. 21), Oxford, Oxford University Press
- Ashley, J. & Devis T. (1992) Death certification from the point of view of the epidemiologist. *Population Trends*, **67**, 22-28
- Baldwin, J.A., Acheson, E.D. & Graham, W.J. (eds.) (1987) *Textbook of Medical Record Linkage*, Oxford, Oxford University Press
- Balzi, D., Buiatti, E., Geddes, M., Khlát, M., Masuyer, E. & Parkin, D.M. (1993) Summary of the results by site. In: Geddes, M., Parkin, D.M., Khlát, M., Balzi, D. & Buiatti, E. (eds.), *Cancer in Italian Migrant Populations* (IARC Scientific Publications No. 123), Lyon, International Agency for Research on Cancer, pp. 193-284
- Bassett, M.T., Chokunonga, E., Mauchaza, B., Levy, L., Ferlay, J. & Parkin, D.M. (1995) Cancer in the African population of Harare, Zimbabwe, 1990-1992. *Int. J. Cancer*, **63**, 29-36
- Beaglehole, R., Bonita, R. & Kjellström, T. (1993) *Basic Epidemiology*, Geneva, World Health Organization
- Beasley, R.P., Hwang, L.-Y., Lin, C.-C. & Chien, C.-S. (1981) Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. *Lancet*, **ii**, 1129-1133
- Belcher, D.W., Wurapa, F.K., Neumann, A.K. & Lourie, I.M. (1976) A household morbidity survey in rural Africa. *Int. J. Epidemiol.*, **5**, 113-120
- Beral, V. (1974) Cancer of the cervix: a sexually transmitted infection? *Lancet*, **1**, 1037-1040
- Beral, V., Chilvers, C. & Fraser, P. (1978) Does pregnancy protect against ovarian cancer? *Lancet*, **1**, 1083-1087
- Beral, V., Chilvers, C. & Fraser, P. (1979) On the estimation of relative risk from vital statistical data. *J. Epidemiol. Community Health*, **33**, 159-162
- Bergström, R., Adami, H.-O., Möhner, M., Zatonski, W., Storm, H., Ekblom, A., Tretli, S., Teppo, L., Akre, O. & Hakulinen, T. (1996) Increase in testicular cancer incidence in six European countries: a birth cohort phenomenon. *J. Natl. Cancer Inst.*, **88**, 727-733
- Bland, M. (1987) *An Introduction to Medical Statistics*, Oxford, Oxford University Press
- Bland, J.M. & Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **1**, 307-310
- Bleehen, N.M. & Stenning, S.P. (The Medical Research Council Brain Tumour Working Party) (1991) A Medical Research Council trial of two radiotherapy doses in the treatment of grades 3 and 4 astrocytoma. *Br. J. Cancer*, **64**, 769-774
- Bone, M.R., Bebbington, A.C., Jagger, C., Morgan, K. & Nicolaas, G. (1995) *Health Expectancy and its Uses*, London, H.M.S.O.
- Bosch, F.X., Muñoz, N., de Sanjosé, S., Izarzugaza, I., Gili, M., Viladiu, P., Tormo, M.J., Moreo, P., Ascunce, N., Gonzalez, L.C., Tafur, L., Kaldor, J.M., Guerrero, E., Aristizabal, N., Santamaria, M., Alonso de Ruiz, P. & Shah, K. (1992) Risk factors for cervical cancer in Colombia and Spain. *Int. J. Cancer*, **52**, 750-758
- Bouchardy, C., Parkin, D.M., Khlát, M., Mirra, A.P., Kogevinas, M., de Lima, F.D. & de Cravalho Ferreira, C.E. (1993) Education and mortality from cancer in São Paulo, Brazil. *Ann. Epidemiol.*, **3**, 64-70
- Bouwes Bavinck, J.N., Vermeer, B.J., van der Woude, F.J., Vandenbroucke, J.P., Schreuder, G.M.Th., Thorogood, J., Persijn, G.G. & Claas, F.H.J. (1991) Relation between skin cancer and HLA antigens in renal-transplant recipients. *New Engl. J. Med.*, **325**, 843-848
- Breslow, N.E. & Day, N.E. (1980) *Statistical Methods in Cancer Research*. Volume I. *The Analysis of Case-Control Studies* (IARC Sci-

- tific Publications No. 32), Lyon, International Agency for Research on Cancer
- Breslow, N.E. & Day, N.E. (1987) *Statistical Methods in Cancer Research*. Volume II. *The Design and Analysis of Cohort Studies* (IARC Scientific Publications No. 82), Lyon, International Agency for Research on Cancer
- Breslow, N.E. & Enstrom, J.E. (1974) Geographic correlations between cancer mortality rates and alcohol-tobacco consumption in the United States. *J. Natl. Cancer Inst.*, **53**, 631-639
- Broders, A.C. (1920) Squamous-cell epithelioma of the lip. A study of five hundred and thirty-seven cases. *J. Am. Med. Assoc.*, **74**, 656-664
- Brothwell, D. (1967) The evidence for neoplasms. In: Brothwell, D. & Sandison, A.T. (eds.) *Diseases in Antiquity: A Survey of Diseases, Injuries and Surgery in Early Populations*, Springfield, IL, Charles C. Thomas, pp. 320-345
- Buck, C., Llopis, A., Nájera, E. & Terris, M. (1988) *The Challenge of Epidemiology. Issues and Selected Readings*. Pan American Health Organization (PAHO) (Scientific Publication No. 505), Washington, DC, PAHO
- Buell, P. & Dunn, J.E. (1965) Cancer mortality among Japanese Issei and Nisei of California. *Cancer*, **18**, 656-664
- Buiatti, E., Palli, D., Amadori, D., Marubini, E., Puntoni, R., Avellini, C., Bianchi, S., Cipriani, F., Cocco, P., Decarli, A., Vindigni, C. & Blot, W. (1989a) Methodological issues in a multicentric study of gastric cancer and diet in Italy: Study design, data sources and quality controls. *Tumori*, **75**, 410-419
- Buiatti, E., Palli, D., Decarli, A., Amadori, D., Avellini, C., Bianchi, S., Bisemi, R., Cipriani, F., Cocco, P., Giacosa, A., Marubini, E., Puntoni, R., Vindigni, C., Fraumeni Jr, J. & Blot, W. (1989b) A case-control study of gastric cancer and diet in Italy. *Int. J. Cancer*, **44**, 611-616
- Buring, J.E., Bain, C.J. & Ehrmann, R.L. (1986) Conjugated estrogen use and risk of endometrial cancer. *Am. J. Epidemiol.*, **124**, 434-441
- Cameron, H.M. & McGoogan, E. (1981) A prospective study of 1152 hospital autopsies: I. Inaccuracies in death certificates. *J. Pathol.*, **113**, 273-283
- Carpenter, L.M. (1987) Some observations on the healthy worker effect. *Br. J. Ind. Med.*, **44**, 289-291
- Case, R.A.M. & Pearson, J.T. (1954) Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Part II. Further consideration of the role of aniline and of the manufacture of auramine and magenta (fuchsine) as possible causative agents. *Br. J. Ind. Med.*, **11**, 213-216
- Case, R.A.M., Hosker, M.E., McDonald, D.B. & Pearson, J.T. (1954) Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Part I. The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine. *Br. J. Ind. Med.*, **11**, 75-104
- Casley, D.J. & Lury, D.A. (1987) *Data Collection in Developing Countries*, second edition, Oxford, Clarendon Press
- Chen, J., Campbell, T.C., Junkvaio, L. & Peto, R. (1990) *Diet, Life-Style and Mortality in China. A Study of the Characteristics of 65 Chinese Counties*, Oxford, Oxford University Press
- Chlebowski, R.T. & Grosvenor, M. (1994) The scope of nutrition intervention trials with cancer-related endpoints. *Cancer*, **74**, 2734-2738
- Chuang, W.-L., Chang, W.-Y., Lu, S.-N., Su, W.-P., Lin, Z.-Y., Chen, S.-C., Hsieh, M.-Y., Wang, L.-Y., You, S.-L. & Chen, C.-J. (1992) The role of hepatitis B and C viruses in hepatocellular carcinoma in a hepatitis B endemic area. A case-control study. *Cancer*, **69**, 2052-2054
- Cicchetti, D.V. & Feinstein, A.R. (1990) High agreement but low Kappa: II. Resolving the paradoxes. *J. Clin. Epidemiol.*, **43**, 551-558
- CIOMS (Council for International Organizations of Medical Sciences) and WHO (World Health Organisation) (1993) *International Ethical Guidelines for Biomedical Research Involving Human Subjects*, Geneva, CIOMS
- Clarke, E.A. & Anderson, T.W. (1979) Does screening by "Pap" smears help prevent cervical cancer? A case-control study. *Lancet*, **2**, 1-4
- Clayton, D. & Hills, M. (1993) *Statistical Models in Epidemiology*, Oxford, Oxford University Press
- Clayton, D. & Schifflers, E. (1987a) Models for temporal variations in cancer rates. I. Age-period and age-cohort models. *Stat. Med.*, **6**, 449-467
- Clayton, D. & Schifflers, E. (1987b) Models for temporal variations in cancer rates. II. Age-period-cohort models. *Stat. Med.*, **6**, 469-481
- Cole, P & Morrison, A.S. (1978) Basic issues in cancer screening. In: Miller, A.B. (ed.) *Screening in Cancer* (A report of a UICC International Workshop, Toronto, Canada, April 24-27, 1978), Geneva, International Union Against Cancer, pp. 7-39
- Collaborative Group on Hormonal Factors in Breast Cancer (1996) Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*, **347**, 1713-1727
- College of American Pathologists (1965) *Systematized Nomenclature of Pathology (SNOP)*, 1st edition, Skolkie, IL
- College of American Pathologists (1977) *Systematized Nomenclature of Medicine (SNOMED)*, 1st edition, Skolkie, IL
- COMMIT Research Group (1991) Community Intervention Trial for Smoking Cessation (COMMIT): summary of design and intervention. *J. Natl. Cancer Inst.*, **83**, 1620-1628
- Cornfield, J. (1951) A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix. *J. Natl. Cancer Inst.*, **11**, 1269-1275
- Cottler, L.B., Zipp, J.F., Robins, L.N. & Spitznagel, E.L. (1987) Difficult-to-recruit respondents and their effect on prevalence estimates in an epidemiologic survey. *Am. J. Epidemiol.*, **125**, 329-339
- Cox, D.R. & Snell, E.J. (1981) *Applied Statistics: Principles and Examples*, London, Chapman & Hall

- Cramer, D.W. (1974) The role of cervical cytology in the declining morbidity and mortality of cervical cancer. *Cancer*, **34**, 2018–2027
- Dawber, T.R., Meadors, G.F. & Moore, F.E., Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. *Am. J. Public Health*, **41**, 279–286
- Day, N.E. & Boice, J.D., Jr (eds.) (1983) *Second Cancer in Relation to Radiation Treatment for Cervical Cancer* (IARC Scientific Publications No. 52), Lyon, International Agency for Research on Cancer
- De Stavola, B. (transl.) (1987) Statistical facts about cancers on which doctor Rigoni-Stern based his contribution to the Surgeon's subgroup of the IV Congress of the Italian scientists on 23 September 1842. *Stat. Med.*, **6**, 881–884
- Dey, P., Collins, S., Will, S. & Woodman, C.B.J. (1995) Randomised controlled trial assessing effectiveness of health education leaflets in reducing incidence of sunburn. *Br. Med. J.*, **311**, 1062–1063
- Doll, R. (1955) Mortality from lung cancer in asbestos workers. *Br. J. Ind. Med.*, **12**, 81–86
- Doll, R. & Hill, A.B. (1950) Smoking and carcinoma of the lung. Preliminary report. *Br. Med. J.*, **ii**, 739–748
- Doll, R. & Hill, A.B. (1952) A study of the aetiology of carcinoma of the lung. *Br. Med. J.*, **ii**, 1271–1286
- Doll, R. & Hill, A.B. (1954) The mortality of doctors in relation to their smoking habits. A preliminary report. *Br. Med. J.*, **i**, 1451–1455
- Doll, R. & Peto, R. (1976) Mortality in relation to smoking: 20 years' observations on male British doctors. *Br. Med. J.*, **2**, 1525–1536
- Doll, R., Payne, P. & Waterhouse, J.A.H. (1966) *Cancer Incidence in Five Continents*. Vol. 1. Geneva, UICC. Berlin, Springer
- Doll, R., Gray, R., Hafner, B. & Peto, R. (1980) Mortality in relation to smoking: 22 years' observations on female British doctors. *Br. Med. J.*, **280**, 967–971
- Doll, R., Peto, R., Wheatley, K., Gray, R. & Sutherland, I. (1994a) Mortality in relation to smoking: 40 years' observations on male British doctors. *Br. Med. J.*, **309**, 901–911
- Doll, R., Peto, R., Hall, E., Wheatley, K. & Gray, R. (1994b) Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors. *Br. Med. J.*, **309**, 911–918
- Dosemeci, M., Wacholder, S. & Lubin, J.H. (1990) Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am. J. Epidemiol.*, **132**, 746–748
- Douglas, A.J., Omar, R.Z. & Smith, P.G. (1994) Cancer mortality and morbidity among workers at the Sellafield plant of British Nuclear Fuels. *Br. J. Cancer*, **70**, 1232–1243
- Draper, G.J., Birch, J.M., Bithell, J.F., Kinnier Wilson, L.M., Leck, I., Marsden, H.B., Morris Jones, P.H., Stiller, C.A. & Swindell, R., (1982) *Childhood Cancer in Britain: Incidence, Survival and Mortality* (Studies in Medical and Population Subjects No. 37), London, H.M.S.O.
- Enterline, P.E. (1965) Mortality among asbestos products workers in the United States. *Ann. N.Y. Acad. Sci.*, **132**, 156–165
- Estève, J., Benhamou, E. & Raymond, L. (1994) *Statistical Methods in Cancer Research. Volume IV. Descriptive Epidemiology*, (IARC Scientific Publications No. 128). Lyon, International Agency for Research on Cancer
- Farr, W. (1975) *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr* (The History of Medicine Series No. 46), Metuchen, NJ, Scarecrow Press.
- Feinleib, M. (1984) Data bases, data banks and data dredging: the agony and the ecstasy. *J. Chron. Dis.*, **37**, 783–790
- Feinstein, A.R. & Cicchetti, D.V. (1990) High agreement but low Kappa: I. The problems of two paradoxes. *J. Clin. Epidemiol.* **43**, 543–549
- Ferlay, J. (1994) *ICD Conversion Programs for Cancer*, (IARC Technical Report No. 21), Lyon, International Agency for Research on Cancer
- Fisher, R.A. & Yates, F. (1963) *Statistical Tables for Biological, Agricultural and Medical Research*, London, Longman
- Fisher, B., Redmond, C., Fisher, E.R., Bauer, M., Wolmark, N., Wickerham, D.L., Deutsch, M., Montague, E., Margolese, R. & Foster, R. (1985) Ten-year results of a randomised clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *New Engl. J. Med.*, **312**, 674–681
- Flegal, K.M., Brownie, C. & Haas, J.D. (1986) The effects of exposure misclassification on estimates of relative risk. *Am. J. Epidemiol.* **123**, 736–751
- Forman, D., Sitas, F., Newell, D.G., Stacey, A.R., Boreham, J., Peto, R., Campbell, T.C., Li, J. & Chen, J. (1990) Geographic association of *Helicobacter pylori* antibody prevalence and gastric cancer mortality in rural China. *Int. J. Cancer*, **46**, 608–611
- Fukao, A., Komatsu, S., Tsubono, Y., Hisamichi, S., Ohori, H., Kizawa, T., Ohsato, N., Fujino, N., Endo, N. & Iha, M. (1993) *Helicobacter pylori* infection and chronic atrophic gastritis among Japanese blood donors: a cross-sectional study. *Cancer Causes Control*, **4**, 307–312
- Gambia Hepatitis Study Group (1987) The Gambia Hepatitis Intervention Study. *Cancer Res.*, **47**, 5782–5787
- Gardner, M.J. & Altman, D.G. (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br. Med. J.*, **292**, 746–750
- Geser, A., de Thé, G., Lenoir, G., Day, N.E. & Williams, E.H. (1982) Final case reporting from the Ugandan prospective study of the relationship between EBV and Burkitt's lymphoma. *Int. J. Cancer*, **29**, 397–400
- Gilliand, F.D. & Samet, J.M. (1994) Lung cancer. In: Doll, R., Fraumeni, J.F., Jr & Muir, C.S. (eds.) *Trends in Cancer Incidence and Mortality, Cancer Surveys*, Vol. 19/20, New York, Cold Spring Harbor Laboratory Press, pp. 175–195
- Gore, S.M. (1981) Assessing methods – survival. *Br. Med. J.*, **283**, 840–843
- Greenland, S. & Morgenstern, H. (1989) Ecological bias, confounding, and effect modification. *Int. J. Epidemiol.*, **18**, 269–274
- Greenland, S. & Robins, J.M. (1985) Estimation of a common

- effect parameter from sparse follow-up data. *Biometrics*, **41**, 55–68
- Greenland, S. & Robins, J. (1994) Invited commentary: Ecologic studies – biases, misconceptions, and counterexamples. *Am. J. Epidemiol.*, **139**, 747–760
- Greenland, S. & Thomas, D.C. (1982) On the need for the rare disease assumption in case-control studies. *Am. J. Epidemiol.*, **116**, 547–553
- Grulich, A.E., Swerdlow, A.J., dos Santos Silva, I. & Beral, V. (1995) Is the apparent rise in cancer mortality in the elderly real? Analysis of changes in certification and coding of cause of death in England and Wales, 1970–1990. *Int. J. Cancer*, **63**, 164–168
- Gupta, P.C., Bhonsle, R.B., Mehta, F.S. & Pindborg, J.J. (1984) Mortality experience in relation to tobacco chewing and smoking habits from a 10-year follow-up study in Ernakulam District, Kerala. *Int. J. Epidemiol.*, **13**, 184–187
- Gupta, P.C., Mehta, F.S., Pindborg, J.J., Aghi, M.B., Bhonsle, R.B., Daftary, D.K., Murti, P.R., Shah, H.T. & Sinor, P.N. (1986) Intervention study for primary prevention of oral cancer among 36000 Indian tobacco users. *Lancet*, **i**, 1235–1239
- Gupta, P.C., Mehta, F.S., Pindborg, J.J., Daftary, D.K., Aghi, M.B., Bhonsle, R.B. & Murti, P.R. (1990) A primary prevention study of oral cancer among Indian villagers. Eight-year follow-up results. In: Hakama, M., Beral, V., Cullen, J.W., Parkin, D.M. (1990) *Evaluating Effectiveness of Primary Prevention of Cancer* (IARC Scientific Publications No. 103), Lyon, International Agency for Research on Cancer, pp. 149–156
- Haenszel, W., Loveland, D. & Sirken, M.G. (1962) Lung cancer mortality as related to residence and smoking histories. *J. Natl Cancer Inst.*, **28**, 947–1001
- Hakama, M. (1982) Trends in the incidence of cervical cancer in the Nordic Countries. In: Magnus, K. (ed.), *Trends in Cancer Incidence. Causes and Practical Implications*, New York, Hemisphere, pp. 279–292
- Hakama, M., Hakulinen, T., Pukkala, E., Saxén, E. & Teppo, L. (1982) Risk indicators of breast and cervical cancer on ecologic and individual levels. *Am. J. Epidemiol.*, **116**, 990–1000
- Hakama, M., Miller, A.B. & Day, N.E. (eds.) (1986) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, p. 289
- Hakulinen, T. & Pukkala, E. (1981) Future incidence of lung cancer: forecasts based on hypothetical changes in the smoking habits of males. *Int. J. Epidemiol.*, **10**, 233–240
- Heasman, M.A. & Clarke, J.A. (1979) Medical record linkage in Scotland. *Health Bull.*, **37**, 97–103
- Heasman, M.A. & Lipworth, L. (1966) *Accuracy of Certification of Cause of Death. A Report on a Survey Conducted in 1959 in 75 Hospitals of the National Health Service to Obtain Information on the Extent of Agreement Between Clinical and Post-mortem Diagnoses* (Studies on Medical and Population Subjects No. 20), London, H.M.S.O.
- Henderson, M.M. (1995) Nutritional aspects of breast cancer. *Cancer*, **76** (suppl. 10), 2053–2058
- Hennekens, C.H. & Eberlein, K. (Physicians' Health Study Research Group) (1985) A randomized trial of aspirin and β -carotene among U.S. physicians. *Prev. Med.*, **14**, 165–168
- Hennekens, C.H., Buring, J.E., Manson, J.E., Stampfer, M., Rosner, B., Cook, N.R., Belanger, C., LaMotte, F., Gaziano, J.M., Ridker, P.M., Willett, W. & Peto, R. (1996) Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New J. Engl. Med.*, **334**, 1145–1149
- Herbst, A.L., Ulfelder, H. & Poskanzer, D.C. (1971) Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl. J. Med.*, **284**, 878–881
- Hill, A.B. (1965) The environment and disease: association or causation? *Proc. R. Soc. Med.*, **58**, 295–300
- Hoffman, F.L. (1915) *The Mortality From Cancer Throughout the World*, Newark, NJ, Prudential Press
- Howe, G.R., Burch, J.D., Miller, A.B., Morrison, B., Gordon, P., Weldon, L., Chambers, L.W., Fodor, G. & Winsor, G.M. (1977) Artificial sweeteners and human bladder cancer. *Lancet*, **2**, 578–581
- IARC (International Agency for Research on Cancer) (1986) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Vol. 38. *Tobacco Smoking*, Lyon, International Agency for Research on Cancer
- IARC (International Agency for Research on Cancer) (1994a) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Vol. 61. *Schistosomes, Liver Flukes and Helicobacter pylori*. Lyon, International Agency for Research on Cancer
- IARC (International Agency for Research on Cancer) (1994b) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Vol. 59. *Hepatitis Viruses*, Lyon, International Agency for Research on Cancer
- Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. & Skeet, R.G. (eds.) (1991) *Cancer Registration. Principles and Methods* (IARC Scientific Publications No. 95), Lyon, International Agency for Research on Cancer
- Joint Iran-International Agency for Research on Cancer Study Group (1977) Esophageal cancer studies in the Caspian littoral of Iran: results of population studies – a pro-drome. *J. Natl. Cancer. Inst.*, **59**, 1127–1138
- Jolley, D (1993) The glitter of the *t* table. *Lancet*, **342**, 27–29
- Jones, W.H.S. (transl.) (1923) *Hippocrates*, Vol. I, London, William Heinemann, pp. 71–137
- Kaldor, J.M., Day, N.E., Bell, J., Clarke, E.A., Langmark, F., Karjalainen, S., Band, P., Pedersen, D., Choi, W., Blair, V., Henry-Amar, M., Prior, P., Assouline, D., Pompe-Kirn, V., Cartwright, R.A., Koch, M., Arslan, A., Fraser, P., Sutcliffe, S.B., Host, H., Hakama, M. & Stovall, M. (1992) Lung can-

- cer following Hodgkin's disease: a case-control study. *Int. J. Cancer*, **52**, 677-681
- Khlat, M., Vail, A., Parkin, M. & Green, A. (1992) Mortality from melanoma in migrants to Australia: variation by age at arrival and duration of stay. *Am. J. Epidemiol.*, **135**, 1103-1113
- Kirkwood, B.R. (1988) *Essentials of Medical Statistics*, London, Blackwell Science
- Kiviat, N.B., Koutsky, L.A., Critchlow, C.W., Galloway, D.A., Vernon, D.A., Peterson, M.L., McElhose, P.E., Pendras, S.J., Stevens, C.E. & Holmes, K.K. (1990) Comparison of Southern transfer hybridization and dot filter hybridization for detection of cervical human papillomavirus infection with types 6, 11, 16, 18, 31, 33, and 35. *Am. J. Clin. Pathol.*, **94**, 561-565
- Kjaer, S.K., de Villiers, E.-M., Haugaard, B.J., Christensen, R.B., Teisen, C., Møller, K.A., Poll, P., Jensen, H., Vestergaard, B.F., Lynge, E. & Jensen, O.M. (1988) Human papillomavirus, herpes simplex virus and cervical cancer incidence in Greenland and Denmark. A population-based cross-sectional study. *Int. J. Cancer*, **41**, 518-524
- Kubik, A., Parkin, D.M., Khlat, M., Erban, J., Polak, J. & Adamec, M. (1990) Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomised controlled trial on a population of high-risk males in Czechoslovakia. *Int. J. Cancer*, **45**, 26-33
- Kurihara, M., Aoki, K. & Hisamichi, S. (eds.) (1989) *Cancer Mortality Statistics in the World 1950-1985*, Nagoya, University of Nagoya Press
- Läärä, E. (1982) *Development of Cancer Morbidity in Finland up to the Year 2002* [in Finnish] (Health Education Series, Original Reports 3/1982), Helsinki, National Board of Health
- Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174
- Lane-Clayton, J.E. (1926) *A Further Report on Cancer of the Breast, with Special Reference to its Associated Antecedent Conditions* (Reports on Public Health and Medical Subjects No. 32), London, H.M.S.O.
- Lantz, C.A. & Nebenzahl, E. (1996) Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J. Clin. Epidemiol.*, **49**, 431-434
- Last, J.M. (1995) *A Dictionary of Epidemiology*, third edition, Oxford, Oxford University Press
- Lee, H.P., Day, N.E. & Shanmugaratnam, K. (1988) *Trends in Cancer Incidence in Singapore 1968-1982* (IARC Scientific Publications No. 91), Lyon, International Agency for Research on Cancer
- Lee, H.P., Gourley, L., Duffy, S.W., Estève, J., Lee, J. & Day, N.E. (1991) Dietary effects on breast-cancer risk in Singapore. *Lancet*, **337**, 1197-1200
- Levin, M.L., Goldstein, H. & Gerhardt, P.R. (1950) Cancer and tobacco smoking. A preliminary report. *J. Am. Med. Assoc.*, **143**, 336-338
- Ley, C., Bauer, H.M., Reingold, A., Schiffman, M.H., Chambers, J.C., Tashiro, C.J. & Manos, M.M. (1991) Determinants of genital human papillomavirus infection in young women. *J. Natl. Cancer Inst.*, **83**, 997-1003
- Lilienfeld, A.M. & Lilienfeld, D.E. (1979) A century of case-control studies: progress? *J. Chron. Dis.*, **32**, 5-13
- Lippman, S.M., Lee, J.S., Lotan, R., Hittelman, W., Wargovich, M.J. & Hong, W.K. (1990) Biomarkers as intermediate end points in chemoprevention trials. *J. Natl. Cancer Inst.*, **82**, 555-560
- Long, E.R. (1928) *A History of Pathology*, London, Baillière, Tindall & Cox
- MacLennan, R. (1991) Items of patient information which may be collected by registries. In: Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. & Skeet, R.G. (eds.) *Cancer Registration: Principles and Methods* (IARC Scientific Publications, No. 95), Lyon, International Agency for Research on Cancer, pp. 43-63
- MacMahon, B., Cole, P., Lin, T.M., Lowe, C.R., Mirra, A.P., Rav-nihar, B., Salber, E.J., Valaoras, V.G. & Yuasa, S. (1970) Age at first birth and breast cancer risk. *Bull. World Health Organ.*, **43**, 209-221
- Malaty, H.M., Kim, J.G., Kim, S.D. & Graham, D.Y. (1996) Prevalence of *Helicobacter pylori* infection in Korean children: inverse relation to socioeconomic status despite a uniformly high prevalence in adults. *Am. J. Epidemiol.*, **143**, 257-262
- Mantel, N. & Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*, **22**, 719-748
- Mayans, M.V., Hall, A.J., Inskip, H.M., Lindsay, S.W., Chotard, J., Mendy, M. & Whittle, H.C. (1994) Do bedbugs transmit hepatitis B? *Lancet*, **343**, 761-763
- McMichael, A.J. (1976) Standardized mortality ratios and the "healthy worker effect": scratching beneath the surface. *J. Occup. Med.*, **18**, 165-168
- McMichael, A.J., Spirtas, R. & Kupper, L.L. (1974) An epidemiologic study of mortality within a cohort of rubber workers, 1964-72. *J. Occup. Med.*, **16**, 458-464
- Meirik, O., Lund, E., Adami, H.-O., Bergström, R., Christoffersen, T. & Bergsjö, P. (1986) Oral contraceptive use and breast cancer in young women. A joint national case-control study in Sweden and Norway. *Lancet*, **2**, 650-654
- Miettinen, O. (1976) Estimability and estimation in case-referent studies. *Am. J. Epidemiol.*, **103**, 226-235
- Miettinen, O.S. & Wang, J.-D. (1981) An alternative to the proportionate mortality ratio. *Am. J. Epidemiol.*, **114**, 144-148
- Møller, H., Knudsen, L.B. & Lynge, E. (1994) Risk of testicular cancer after vasectomy: cohort study of over 73 000 men. *Br. Med. J.*, **309**, 295-299
- Morrison, A.S., Buring, J.E., Verhoek, W.G., Aoki, K., Leck, I., Ohno, Y. & Obata, K. (1984) An international study of smoking and bladder cancer. *J. Urology*, **131**, 650-654
- Mould, R.F. (1983) *Cancer Statistics* (Medical Sciences Series), Bristol, Adam Hilger
- Mueller, N., Evans, A., Harris, N.L., Comstock, G.W., Jellum, E., Magnus, K., Orentreich, N., Polk, B.F. & Vogelstein, J. (1989) Hodgkin's disease and Epstein-Barr virus. Altered an-

- tibody pattern before diagnosis. *New Engl. J. Med.*, **320**, 689–695
- Muir, C.S. & Percy, C. (1991) Classification and coding of neoplasms. In: Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. & Skeet, R.G. (eds.) *Cancer Registration. Principles and Methods* (IARC Scientific Publications, No. 95), Lyon, International Agency for Research on Cancer, pp. 64–81
- Muir, C., Waterhouse, J., Mack, T., Powell, J. & Whelan, S. (eds.) (1987) *Cancer Incidence in Five Continents*. Vol. V (IARC Scientific Publications, No. 88), Lyon, International Agency for Research on Cancer
- Muñoz, N. & Day, N.E. (1996) Oesophageal cancer. In: Schottenfeld, D., Fraumeni Jr., J.F. (eds.) *Cancer Epidemiology and Prevention*, second edition. Oxford, Oxford University Press, pp. 681–706
- Muñoz, N., Wahrendorf, J., Bang, L.J., Crespi, M., Thurnham, D.I., Day, N.E., Ji, Z.H., Grassi, A., Yan, L.W., Lin, L.G., Quan, L.Y., Yun, Z.C., Fang, Z.S., Yao, L.J., Correa, P., O'Connor, G.T. & Bosch, X. (1985) No effect of riboflavine, retinol, and zinc on prevalence of precancerous lesions of oesophagus. Randomised double-blind intervention study in high-risk population of China. *Lancet*, **2**, 111–114
- Muñoz, N., Bosch, F.X., de Sanjosé, S., Tafur, L., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C., Ascunce, N., Gonzalez, L.C., Kaldor, J.M., Guerrero, E., Lőrincz, A., Santamaria, M., Alonso de Ruiz, P., Aristizabal, N. & Shah, K. (1992a) The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Colombia and Spain. *Int. J. Cancer*, **52**, 743–749
- Muñoz, N., Bosch, F.X., Shah, K.V. & Meheus, A. (eds.) (1992b) *The Epidemiology of Cervical Cancer and Human Papillomavirus* (IARC Scientific Publications, No. 119), Lyon, International Agency for Research on Cancer
- Newton, R., Ferlay, J., Reeves, G., Beral, V. & Parkin, D.M. (1996) Effect of ambient solar ultraviolet radiation on incidence of squamous-cell carcinoma of the eye. *Lancet*, **347**, 1450–1451
- Nomura, A., Stemmermann, G.N., Chyou, P.H., Kato, I., Perez-Perez, G.I. & Blaser, M.J. (1991) *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. *New Engl. J. Med.*, **325**, 1132–1136
- OPCS (Office of Population Censuses and Surveys) (1970) *Classification of Occupations, 1970*. London, H.M.S.O.
- OPCS (Office of Population Censuses and Surveys) (1978) *Occupational Mortality. The Registrar General's Decennial Supplement for England and Wales, 1970–72* (Series DS, No. 1), London, H.M.S.O.
- OPCS (Office of Population Censuses and Surveys) (1981) *Cancer Statistics. Incidence, Survival and Mortality in England and Wales* (Studies on Medical and Population Subjects, No. 43), London, H.M.S.O.
- OPCS (Office of Population Censuses and Surveys) (1983a, 1983b, 1985a, 1985b, 1986, 1988, 1990, 1991, 1993, 1994a, 1994b) *Cancer Statistics. Registrations* (Series MB1 Nos. 11–16, 18–22), London, H.M.S.O.
- OPCS (Office of Population Censuses and Surveys) (1995) *Mortality Statistics. Review of the Registrar General on Deaths by Cause, Sex and Age, in England and Wales, 1993* (Series DH2 No. 20), London, H.M.S.O.
- Omenn, G.S., Goodman, G.E., Thornquist, M.D., Balmes, J., Cullen, M.R., Glass, A., Keogh, J.P., Meyskens, F.L., Valanis, B., Williams, J.H., Barnhart, S. & Hammar, S. (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New Engl. J. Med.*, **334**, 1150–1155
- Orr, I.M. (1933) Oral cancer in betel nut chewers in Travancore. Its aetiology, pathology and treatment. *Lancet*, **ii**, 575–580
- Parkin, D.M. & Hakulinen, T. (1991) Analysis of survival. In: Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. & Skeet, R.G. (eds.) *Cancer Registration: Principles and Methods* (IARC Scientific Publications No. 95), Lyon, International Agency for Research on Cancer, pp. 159–176
- Parkin, D.M. & Khlat, M. (1996) Studies of cancer in migrants: rationale and methodology. *Eur. J. Cancer*, **32A**, 761–771
- Parkin, D.M., Muir, C.S., Whelan, S.L., Gao, Y.T., Ferlay, J. & Powell, J. (1992) *Cancer Incidence in Five Continents*. Vol. VI (IARC Scientific Publications No. 120), Lyon, International Agency for Research on Cancer
- Parkin, D.M., Pisani, P. & Ferlay, J. (1993) Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int. J. Cancer*, **54**, 594–606
- Parkin, D.M., Chen, V.W., Ferlay, J., Galceran, J., Storm, H.H. & Whelan, S.L. (1994) *Comparability and Quality Control in Cancer Registration* (IARC Technical Reports No. 19), Lyon, International Agency for Research on Cancer
- Parkin, D.M., Whelan, S.L., Ferlay, J., Raymond, L. & Young, J. (1997) *Cancer Incidence in Five Continents*. Vol. VII (IARC Scientific Publications No. 143), Lyon, International Agency for Research on Cancer
- Pearce, N., Matos, E., Vainio, H., Boffetta, P. & Kogevinas, M. (eds) (1994) *Occupational Cancer in Developing Countries* (IARC Scientific Publications No. 129), Lyon, International Agency for Research on Cancer
- Percy, C. L. (ed.) (1980) *Conversion of Malignant Neoplasms by Topography and Morphology from the International Classification of Diseases for Oncology (ICD-O) to Chapter II, Malignant Neoplasms*, International Classification of Diseases (ICD-8), 1965 (NIH Publication No. 80–2136), Bethesda, MD, US Department of Health, Education and Welfare
- Percy, C. L. (ed.) (1981) *Conversion of Neoplasm Section, 8th Revision of International Classification of Diseases (1965): and 8th revision International Classification of Diseases Adapted for Use in the United States, to Neoplasm Section, 9th Revision of International Classification of Diseases (1975)* (NIH Publication No. 83–2448), Bethesda, MD, US Department of Health and Human Services
- Percy, C.L. (ed.) (1983) *Conversion of Neoplasm Section, 9th Revision of International Classification of Diseases (1975): to Neo-*

- plasm Section, 8th Revision of International Classification of Diseases (1965) and 8th Revision of International Classification of Diseases, Adapted for Use in the United States (NIH Publication No. 83-2638), Bethesda, MD, US Department of Health and Human Services
- Percy, C.L., Berg, J.W. & Thomas, L.B. (eds.) (1968) *Manual of Tumour Nomenclature and Coding*, second edition. Washington, DC, American Cancer Society
- Percy, C.L. & van Holten, V. (eds.) (1979) *Conversion of Neoplasms by Topography and Morphology from the International Classification of Diseases for Oncology (ICD-O): to Chapter II, Neoplasms, 9th Revision of the International Classification of Diseases* (NIH Publication No. 80-2007), Bethesda, MD, US Department of Health, Education and Welfare, Public Health Service
- Percy, C., van Holten, V. & Muir, C. (eds.) (1990) *International Classification of Diseases for Oncology (ICD-O)*, second edition. Geneva, World Health Organization
- Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. & Smith, P.G. (1976) Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br. J. Cancer*, **34**, 585-612
- Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. & Smith, P.G. (1977) Design and analysis of randomised clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br. J. Cancer*, **35**, 1-39
- Pike, M.C., Henderson, B.E., Krailo, M.D., Duke, A. & Roy, S. (1983) Breast cancer in young women and use of oral contraceptives: possible modifying effect of formulation and age at use. *Lancet*, **2**, 926-930
- Pocock, S.J. (1983) *Clinical Trials: A Practical Approach*, Chichester, John Wiley
- Polednak, A.P., Stehney, A.F. & Rowland, R.E. (1978) Mortality among women first employed before 1930 in the U.S. radium dial-painting industry. A group ascertained from employment lists. *Am. J. Epidemiol.*, **107**, 179-195
- Potosky, A.L., Kessler, L., Gridley, G., Brown, C.C. & Horm, J.W. (1990) Rise in prostatic cancer incidence associated with increased use of transurethral resection. *J. Natl. Cancer Inst.*, **82**, 1624-1628
- Pott, P. (1775) *Chirurgical Observations*. Vol. 3, London, Hawes, L., Clark, W, Collins, R., pp. 177-183.
- Pottern, L.M., Stone, B.J., Day, N.E., Pickle, L.W. & Fraumeni, J.F., Jr (1980) Thyroid cancer in Connecticut, 1935-1975: an analysis by cell type. *Am. J. Epidemiol.*, **112**, 764-774
- Powles, T.J., Jones, A.L., Ashley, S.E., O'Brien, M.E.R., Tidy, V.A., Treleavan, J., Cosgrove, D., Nash, A.G., Sacks, N., Baum, M., McKinna, J.A. & Davey, J.B. (1994) The Royal Marsden Hospital pilot tamoxifen chemoprevention trial. *Breast Cancer Res. Treat.*, **31**, 73-82
- Prince, A.M., Szmuness, W., Michon, J., Demaille, J., Diebolt, G., Linhard, J., Quenum, C. & Sankale, M. (1975) A case/control study of the association between primary liver cancer and hepatitis B infection in Senegal. *Int. J. Cancer*, **16**, 376-383
- Puffer, R.R. & Griffith, G. W. (1967) *Patterns of Urban Mortality. Report of the Inter-American Investigation of Mortality* (Scientific Publication No. 151), Washington, DC, Pan American Health Organization
- Qian G.-S., Ross, R.K, Yu, M.C., Yuan, J.-M., Gao, Y.-T., Henderson, B.E., Wogan, G.N. & Groopman, J.D. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol. Biomarkers Prev.*, **3**, 3-10
- Rodrigues, L. & Kirkwood, B.R. (1990) Case-control designs in the study of common diseases: Updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int. J. Epidemiol.*, **19**, 205-213
- Romieu, I., Willett, W.C., Colditz, G.A., Stampfer, M.J., Rosner, B., Hennekens, C.H. & Speizer, F.E. (1989) Prospective study of oral contraceptive use and risk of breast cancer in women. *J. Natl. Cancer Inst.*, **81**, 1313-1321
- Rose, G. (1985) Sick individuals and sick populations. *Int. J. Epidemiol.*, **14**, 32-38
- Rose, G. & Colwell, L. (1992) Randomised controlled trial of anti-smoking advice: final (20 year) results. *J. Epidemiol. Comm. Health*, **46**, 75-77
- Ross, D.A. & Vaughan, J.P. (1986) Health interview surveys in developing countries: a methodological review. *Stud. Fam. Plann.*, **17**, 78-94
- Ross, R.K., Yuan, J.-M., Yu, M.C., Wogan, G.N., Qian, G.-S., Tu, J.-T., Groopman, J.D., Gao, Y.-T. & Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, **339**, 943-946
- Sackett, D.L. (1979) Bias in analytic research. *J. Chron. Dis.*, **32**, 51-63
- Schatzkin, A., Freedman, L.S., Schiffman, M.H. & Dawsey, S.M. (1990) Validation of intermediate end points in cancer research. *J. Natl. Cancer Inst.*, **82**, 1746-1752
- Schrek, R. & Lenowitz, H. (1947) Etiologic factors in carcinoma of the penis. *Cancer Res.*, **7**, 180-187
- Schrek, R., Baker L.A., Ballard, G.P. & Dolgoff, S. (1950) Tobacco smoking as an etiologic factor in disease. I. Cancer. *Cancer Res.*, **10**, 49-58
- Schwartz, D. & Lellouch, J. (1967) Explanatory and pragmatic attitudes in therapeutical trials. *J. Chron. Dis.*, **20**, 637-648
- Scott, C. & Harpham, T. (1987) Sample design. In: Cleland, J. & Scott, C. (eds.), *The World Fertility Survey. An Assessment*, Oxford, Oxford University Press, pp. 309-324
- Severson, R.K. (1990) Have transurethral resections contributed to the increasing incidence of prostatic cancer? *J. Natl. Cancer Inst.*, **82**, 1597-1598
- Shapiro, S. (1977) Evidence on screening for breast cancer from a randomised trial. *Cancer*, **39** (suppl. 6), 2772-2782
- Shapiro, S., Strax, P. & Venet, L. (1971) Periodic breast cancer screening in reducing mortality from breast cancer. *J. Am. Med. Assoc.*, **215**, 1777-1785

- Shapiro, S., Goldberg, J.D. & Hutchison, G.B. (1974) Lead time in breast cancer detection and implications for periodicity of screening. *Am. J. Epidemiol.*, **100**, 357–366
- Shimizu, Y., Schull, W.J. & Kato, H. (1990) Cancer risk among atomic bomb survivors. The RERF Life Span Study. *J. Am. Med. Assoc.*, **264**, 601–604
- Shimizu H., Ross, R.K., Bernstein, L., Yatani, R., Henderson, B.E. & Mack, T.M. (1991) Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. *Br. J. Cancer*, **63**, 963–966
- Shu, X.O., Nesbit, M.E., Buckley, J.D., Krailo, M.D. & Robison, L.L. (1995) An exploratory analysis of risk factors for childhood malignant germ-cell tumors: report from the Childrens Cancer Group (Canada, United States) *Cancer Causes Control*, **6**, 187–198
- Smith, P.G. & Day, N.E. (1984) The design of case-control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.*, **13**, 356–365
- Smith, P.G. & Douglas, A.J. (1986) Mortality of workers at the Sellafield plant of British Nuclear Fuels. *Br. Med. J.*, **293**, 845–854
- Smith, P.G. & Morrow, R.H. (eds.) (1996) *Field Trials of Health Interventions in Developing Countries: A Toolbox*, second edition. London, Macmillan Education
- Smith, D.C., Prentice, R., Thompson, D.J. & Herrmann, W.L. (1975) Association of exogenous estrogen and endometrial carcinoma. *New Engl. J. Med.*, **293**, 1164–1167
- Smith, P.G., Rodrigues, L.C. & Fine, P.E.M. (1984) Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *Int. J. Epidemiol.*, **13**, 87–93
- Smith, D.B., Newlands, E.S., Rustin, G.J.S., Begent, R.H.J., Howells, N., McQuade, B. & Bagshawe, K.D. (1991) Comparison of ondansetron and ondansetron plus dexamethasone as antiemetic prophylaxis during cisplatin-containing chemotherapy. *Lancet*, **338**, 487–490
- Sparén, P. (1996) Early Detection and Screening for Cancer of the Cervix in Sweden during the 20th Century (Doctoral thesis. University of Uppsala, Sweden), Stockholm, Almqvist & Wiksell
- Sriamporn, S., Black, R.J., Sankaranarayanan, R., Kamsa-AD, S., Parkin, D.M. & Vatanasapt, V. (1995) Cancer survival in Khon Kaen Province, Thailand. *Int. J. Cancer*, **61**, 296–300
- Stadel, B.V., Rubin, G.L., Webster, L.A., Schlesselman, J.J. & Wingo, P.A. (1985) Oral contraceptives and breast cancer in young women. *Lancet*, **2**, 970–973
- Stampfer, M.J., Buring, J.E., Willet, W., Rosner, B., Eberlein, K. & Hennekens, C.H. (1985) The 2x2 factorial design: its application to a randomized trial of aspirin and carotene in U.S. physicians. *Stat. Med.*, **4**, 111–116
- Strader, C.H., Weiss, N.S. & Daling, J.R. (1988) Vasectomy and the incidence of testicular cancer. *Am. J. Epidemiol.*, **128**, 56–63
- Swerdlow, A.J. (1989) Interpretation of England and Wales cancer mortality data: the effect of enquiries to certifiers for further information. *Br. J. Cancer*, **59**, 787–791
- Swerdlow, A.J., Marmot, M.G., Grulich, A.E. & Head, J. (1995) Cancer mortality in Indian and British ethnic immigrants from the Indian subcontinent to England and Wales. *Br. J. Cancer*, **72**, 1312–1319
- Tabár, L., Fagerberg, C.J.G., Gad, A., Baldetorp, L. Holmberg, L.H., Gröntoft, O., Ljungquist, U., Lundström, B., Månson, J.C., Eklund, G., Day, N.E. & Pettersson, F. (1985) Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet*, **1**, 829–832
- Thames Cancer Registry (1995) *Cancer in South East England 1992. Cancer Incidence, Prevalence, Survival and Treatment for Residents of the District Health Authorities in South East England*, Sutton, Thames Cancer Registry
- Tomášek, L., Darby, S.C., Swerdlow, A.J., Plaček, V. & Kunz, E. (1993) Radon exposure and cancers other than lung cancer among uranium miners in West Bohemia. *Lancet*, **341**, 919–923
- Tomášek, L., Swerdlow, A.J., Darby, S.C., Plaček, V. & Kunz, E. (1994) Mortality in uranium miners in West Bohemia: a long term cohort study. *Occup. Environ. Med.*, **51**, 308–315
- Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. & Pearce, N. (1997) *Application of Biomarkers in Cancer Epidemiology* (IARC Scientific Publications No. 142), Lyon, International Agency for Research on Cancer
- Trichopoulos, D., Tabor, E., Gerety, R.J., Xirouchaki, E., Sparros, L., Muñoz, N. & Linsell, C.A. (1978) Hepatitis B and primary hepatocellular carcinoma in a European population. *Lancet*, **ii**, 1217–1219
- Tufte, E.R. (1983) *The Visual Display of Quantitative Information*, Cheshire, CT, Graphic Press
- Tuyns, A.J. & Massé, L.M.F. (1973) Mortality from cancer of the oesophagus in Brittany. *Int. J. Epidemiol.*, **2**, 241–245
- Tuyns, A.J. & Vernhes, J.C. (1981) La mortalité par cancer de l'oesophage dans les départements du Calvados et de l'Orne. *Gastroenterol. Clin. Biol.*, **5**, 257–265
- Tuyns, A.J., Pequignot, G. & Jensen, O.M. (1977) Le cancer de l'oesophage en Ille-et-Vilaine en fonction des niveaux de consommation d'alcool et de tabac. Des risques qui se multiplient. *Bull. Cancer*, **64**, 45–60
- Tuyns, A.J., Hu, M.X. & Pequignot, G. (1983) Alcohol consumption patterns in the department of Calvados (France) *Rev. Epidemiol. Santé Publique*, **31**, 179–197
- United Nations (1987) *Fertility Behaviour in the Context of Development* (Population Studies No. 100), New York, United Nations, Department of International Economic and Social Affairs
- Uppsala-Örebro Breast Cancer Study Group (1990) Sector resection with or without postoperative radiotherapy for Stage I breast cancer: A randomized trial. *J. Natl. Cancer Inst.*, **82**, 277–282
- US Department of Health, Education, and Welfare (1969) *Use of Tobacco: Practices, Attitudes, Knowledge, and Beliefs. United States - Fall 1964 and Spring 1966*, Washington, DC, US Department of Health, Education, and Welfare
- US Department of Health and Human Services (1991) *Smoking and*

- Tobacco Control Monographs 1. Strategies to Control Tobacco Use in the United States: A Blueprint for Public Health Action in the 1990's* (NIH Publication No. 92-3316), Bethesda, MD, US Department of Health and Human Services, US Government Printing Office
- US Surgeon General (1982) *The Health Consequences of Smoking: Cancer*, Bethesda, MD, US Department of Health and Human Services
- Veronesi, U., Saccozzi, R., Del Vecchio, M., Banfi, A., Clemente, C., De Lena, M., Gallus, G., Greco, M., Luini, A., Marubini, E., Muscolino, G., Rilke, F., Salvadori, B., Zecchini, A. & Zucali, R. (1981) Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. *New Engl. J. Med.*, **305**, 6-11
- Vineis, P. & Simonato, L. (1986) Estimates of the proportion of bladder cancers attributable to occupation. *Scand. J. Work Environ. Health*, **12**, 55-60
- Vizcaino, A.P., Parkin, D.M. & Skinner, M.E.G. (1995) Risk factors associated with oesophageal cancer in Bulawayo, Zimbabwe. *Br. J. Cancer*, **72**, 769-773
- Wagner, G. (1991) History of cancer registration. In: Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S. & Skeet, R.G. (eds.) *Cancer Registration: Principles and Methods* (IARC Scientific Publications, No. 95), Lyon, International Agency for Research on Cancer, pp. 3-6
- Waldron, H.A. (1983) A brief history of scrotal cancer. *Br. J. Ind. Med.*, **40**, 390-401
- Walter, S.D. (1991a) The ecologic method in the study of environmental health. I. Overview of the method. *Environ. Health Perspect.*, **94**, 61-65
- Walter, S.D. (1991b) The ecologic method in the study of environmental health. II. Methodologic issues and feasibility. *Environ. Health Perspect.*, **94**, 67-73
- Waterhouse, J.A.H., Doll, R. & Muir, C.S. (eds.) (1970) *Cancer Incidence in Five Continents*. Vol. II, Berlin, Springer
- Waterhouse, J.A.H., Muir, C.S. & Correa, P. (eds.) (1976) *Cancer Incidence in Five Continents*. Vol. III (IARC Scientific Publications No. 15), Lyon, International Agency for Research on Cancer
- Waterhouse, J.A.H., Muir, C.S., Shanmugaratnam, K. & Powell, J. (eds.) (1982) *Cancer Incidence in Five Continents*. Vol. IV (IARC Scientific Publications No. 42), Lyon, International Agency for Research on Cancer
- Weinstock, M.A., Colditz, G.A., Willett, W.C., Stampfer, M.J., Rosner, B. & Speizer, F.E. (1991) Recall (report) bias and reliability in the retrospective assessment of melanoma risk. *Am. J. Epidemiol.*, **133**, 240-245
- Whelan, S.L., Parkin, D.M. & Masuyer, E. (eds.) (1990) *Patterns of Cancer in Five Continents* (IARC Scientific Publications No. 102), Lyon, International Agency for Research on Cancer
- WHO (World Health Organization) (1956) *Statistical Code for Human Tumours* (WHO/HS/CANC, 24.1 and 24.2, August 15, 1956), Geneva, WHO
- WHO (World Health Organization) (1957) *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*, seventh revision, Geneva, WHO
- WHO (World Health Organization) (1967) *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*, eighth revision, Geneva, WHO
- WHO (World Health Organization) (1976) *International Classification of Diseases for Oncology (ICD-O)*, first edition, Geneva, WHO
- WHO (World Health Organization) (1977) *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*, ninth revision, Geneva, WHO
- WHO (World Health Organization) (1983) *World Health Statistics Annual, 1983*, Geneva, WHO
- WHO (World Health Organization) (1986) *World Health Statistics Annual, 1986*, Geneva, WHO
- WHO (World Health Organization) (1992) *International Statistical Classification of Diseases and Related Health Problems*, tenth revision (ICD-10), Geneva, WHO
- WHO Collaborative Study of Neoplasia and Steroid Contraceptives (1989) Combined oral contraceptives and liver cancer. *Int. J. Cancer*, **43**, 254-259
- Willett, W.C., Stampfer, M.J., Colditz, G.A., Rosner, B.A., Hennekens, C.H. & Speizer, F.E. (1987) Dietary fat and the risk of breast cancer. *New Engl. J. Med.*, **316**, 22-28
- Wright, W.C. (transl.) (1964) *Diseases of Workers*, by Ramazzini, B. (The History of Medicine Series No. 23), New York, Hafner
- Wynder, E.L. & Graham, E.A. (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. A study of six hundred and eighty-four proved cases. *J. Am. Med. Assoc.*, **143**, 329-336
- Ziel, H.K. & Finkle, W.D. (1975) Increased risk of endometrial carcinoma among users of conjugated estrogens. *New Engl. J. Med.*, **293**, 1167-1170

Índice de materias

- Ajuste
 - por edad 75-83
 - por factores de confusión 328-333, 336-340, 384
- Aleatorio
 - Distribución 149, 150
 - Números 171
 - Muestreo 109, 229-230
- Aleatorización 91, 149, 312, 398
 - de pares apareados 153
 - estratificada 152, 232
 - por bloques 152
 - restringida (limitada) 152
 - simple 151, 229
- Análisis
 - de estudios apareados 219
 - de estudios basados en datos rutinarios 257-259, 266-271
 - de estudios caso-control 214-219
 - de estudios de cohorte 186-192
 - de nacimiento 73
 - de estudios ecológicos 266
 - de estudios transversales 237-238
 - de la supervivencia 279-293, 424-426
 - Paquetes estadísticos 292
 - de regresión 267, 340-345
 - de subgrupos 167
 - del tratamiento aleatorizado 161
 - final 163
 - por intención de tratar 161-162
 - seriado 162
- Apareados
 - Diseño de grupos 153
 - Estudios 219
- Apareamiento
 - individual 211, 219
 - excesivo 208, 212-213
 - por frecuencia 213
- Azar 109-124, 314
- Biológica
 - Dosis eficaz 14
 - Muestra 182, 193, 444
- Plausibilidad 17-18
- Calidad
 - Control, laboratorio 32, 444
 - de los datos 33-35, 261, 415
 - Indicadores 34
- Caso
 - incidente 204
 - Definición 61, 201, 279
 - prevalente 204
- Caso-cohorte
 - Estudios 210
- Causa de muerte 33, 262
 - Clasificación errónea 33
- Causalidad 315-318
 - inversa 224
- Censo 70, 262
- Censura 283-285
- Certificado de defunción 33, 243, 257
- Ciego (*véase* Técnicas de ciego)
- Clasificación
 - de la enfermedad 43-46, 415
 - Clasificación Internacional de Enfermedades 6, 43-46
 - Clasificación Internacional de Enfermedades para Oncología 44, 415
 - errónea
 - de la causa de muerte 33-34
 - de la enfermedad o resultado 201, 206, 309-310.
 - de la exposición 301-302
 - diferencial 31-32, 305-310
 - no diferencial 31, 303-304
- Coefficiente de correlación 267
- Comité
 - de monitorización de datos 162
 - ético 147, 433
- Confusión 11, 311-314, 374
 - Control 312, 325-346
 - en estudios basados en datos de rutina 263
 - en estudios caso-control 211
 - en estudios de cohorte 183, 194
 - en estudios ecológicos 272

- en estudios de intervención 150
 - residual 313
- Consentimiento 146
- Consideraciones éticas 89, 141, 144-146, 433
- Contaminación 158
- Controles
 - del vecindario 208
 - en estudios caso-control 95, 206-214, 221
 - en estudios hospitalarios 208, 298-299
 - Esquemas de muestreo 210
 - Número 209
- Cribaje 378, 387-403
 - Efectividad 395-398
 - Evaluación 394-402, 424
 - Monitorización 255
 - Población 392
- Criterios de inclusión 146
- Cuartiles 52
- Cuestionarios 15
 - administrados por el entrevistador 15
 - auto-administrados 15
 - Diseño 37-41
 - estándar 41
 - Evaluación 40
- Cumplimiento 157
- Curva ROC 26
- Datos
 - basales 157, 164
 - Calidad 33-35, 261, 415
 - categoricos 53
 - Codificación 45
 - Control de calidad 48, 450
 - cualitativos 53
 - cuantitativos 49
 - de incidencia 34, 261
 - de mortalidad 33, 261-262
 - Edición 48, 445
 - Introducción en el ordenador 449
 - para registros del cáncer 412-415
 - Procesamiento 47, 445
 - Recopilación 412-414, 445
 - Reducción 49
 - rutinarios 33-35, 69, 97-101, 184, 243-277
 - Validez 20, 26
- Denominador 62, 69, 262, 411
 - Falta de 83, 417
- Desenmascarar, romper el cegado 161
- Desviación estándar 52
 - de la distribución del muestreo 110
- Determinaciones
 - ambientales 18
 - biológicas 17-18
 - de laboratorio 18, 32, 443
 - de la exposición 13-19, 180-183, 214, 274-275
 - ambientales 18
 - biológicas 17
- Diagnóstico 200, 279
 - Base para el diagnóstico 413
 - Sobrediagnóstico 397
 - Sesgo 221
- Diagramas
 - de barras 53
 - de cajas «*box-and-whisker*» 52
 - de dispersión 265, 267
 - de Lexis 71-72
- Diarios 16
- Diseño
 - de cuestionarios 37-41
 - de grupos apareados 153
 - del estudio 89-107 (*véase también* estudios caso-control, estudios de cohorte, estudios transversales, estudios de intervención, estudios basados en datos rutinarios)
 - factorial 154
- Distribución
 - aleatoria 149, 150
 - de muestreo 109
 - Desviación estándar 110
 - normal 53, 356
- Dosis 14
 - biológica eficaz 14 (*véase también* Exposición)
- Ecológico
 - Estudio 99, 263-274, 399
 - Falacia 271-272
- Edad
 - Ajuste 75-83
 - Tasa específica 71, 75-76
- Efecto
 - adverso 157
 - del trabajador sano 179, 263, 299
 - Modificación 335
- Eficacia de la vacuna 167
- Enfermedad preclínica 377-378
- Enmascaramiento 161
- Ensayos
 - aleatorizados 89-93, 141-170
 - clínicos 91, 141
 - comunitarios 142, 158
 - cruzados 155

- de campo 91-93
- Entrevista 15
- Epidemiología molecular 17
- Error
 - de medida 29-32
 - de muestreo 109-110
 - estándar 110, 113, 125, 355
 - de la media 113
- Especificidad 21, 392
- Estadístico
 - Kappa* 27
 - Poder 175, 202, 209, 355-363
 - Prueba 133-139, 347-351
 - Significación 119-120, 314, 355
- Estándar de oro 22-23
- Estandarización 77, 87
 - directa 77
 - indirecta 79
- Estratificación 70-71, 313, 325-327
- Estudios
 - apareados 219
 - basados en datos rutinarios 97-101, 243-276
 - caso-control 7, 95, 199-224
 - Combinación 320
 - de cohorte 6, 93, 173-198
 - de intervención 89-93, 141-170
 - de migrantes 245-250, 300
 - ecológicos 99, 263-274, 399
 - multicéntricos 214
 - nutricionales 193
 - ocupacionales 175-176, 181, 250, 299, 310
 - observacionales 89-90, 398-399
 - transversales 96, 225-241
- Estudios apareados
 - Análisis 219
 - Diseño de grupos apareados 153
- Estudios basados en datos rutinarios
 - a nivel agregado 263-275
 - a nivel individual 243-263
 - Análisis 257-259, 266-271
- Estudios caso-control
 - Análisis 214-219
 - anidados 192-193, 207
 - de cribaje 401, 405-407
 - Determinación de la exposición 214
 - Fuente de casos 205
 - hospitalarios 205
 - Selección de controles 95-96, 206-214
 - Sesgos 202, 207-209, 214, 296-299
- Estudios de cohorte
 - de cribaje 401
 - históricos 180
 - prospectivos 180
 - Sesgos 173, 179-180, 184-185, 194, 299
- Estudios de intervención
 - de cribaje 398-399
 - de tratamiento del cáncer 403
 - Sesgo 300
- Estudios transversales
 - Análisis 237-238
 - Sesgos 234, 296
 - Recopilación de datos 237
 - Tamaño de muestra 227-237
- Exceso
 - Fracción de exceso 104-105, 379-380, 405
 - de la población 382, 405
 - Riesgo en exceso 103-105, 378-179, 405
 - de la población 382, 405
- Exposición
 - acumulativa 14
 - Clasificación errónea 301-310
 - Determinación 13-19, 180-183, 214, 274-275
 - Determinaciones ambientales 18
 - Determinaciones biológicas 17
 - Fuentes de datos 15-19
 - Reducción 377, 385-387
 - Relación dosis-respuesta / exposición-respuesta 189-191, 317
 - Tasa 14
- Fiabilidad
 - inter-observador 26-29
 - intra-observador 26
- Financiación de proyectos 432-436
- Fracción prevenible 105, 167, 380
- Gráficos 57-58
- Grupos
 - de alto riesgo 402
 - de comparación 176
 - externos 177
 - internos 176
 - múltiples 178
- Hipótesis 200
 - Contraste 115
 - nula 117, 133, 355
 - Prueba 117
- Histograma 50
- Identificadores personales 35

- Incidencia 64
 específica por edad 71, 75-76
 cruda 70
 Datos 34, 261
 del cáncer 34, 261
 Fecha 413
 proporcional 84
 Razón de incidencia estandarizada 80
 Razón de incidencia proporcional 84, 257
 Tasa 66-68
- Interacción 335-336, 374
- Intervalo de confianza 111, 122, 314
 para medidas de aparición de la enfermedad 125-128
 para medidas del efecto 128-132, 347-351
 y cálculo del tamaño del estudio 355
- Kaplan-Meier 287 (*véase también* Análisis de supervivencia)
- Mantel-Haenszel
 con múltiples factores de confusión 336-340
 Odds ratio 328, 336, 347
 Prueba 133, 328-333
 Razón de riesgos 329, 349
 Razón de tasas 331, 350
- Manual de Nomenclatura de Tumores y Codificación (MOT-NAC) 44
- Marcadores
 exposición 18
- Media
 aritmética 50-51
 Error estándar 113
 muestral 113
- Mediana 51
- Medidas
 ambientales 18
 biológicas 17-18
 de laboratorio 18, 32, 443
 de la exposición 13-15, 180-183, 214
 de variables de confusión 183
 del resultado 19-20, 184
 Error 29-32
 Sesgo 161, 184, 194, 301-310
- Meta-análisis 320
- Modelo de riesgos proporcionales de Cox 345
- Mortalidad
 Datos 33, 261-262
 Razón de mortalidad estandarizada 80
 Razón de mortalidad proporcional 84, 257
- Muestra
 aleatoria 109
 biológica 181-182, 193, 442-445
 Cálculo del tamaño 353-375
 de la población 109, 227
 de sangre 193, 444
- Muestreo
 aleatorio 109, 228-233
 estratificado 232
 de controles 210
 Determinaciones ambientales 18
 Distribución 109
 en múltiples etapas 230
 Error 109-110
 Marco 228
 sistemático 230
- Negativa a participar 156, 161, 300
- Nivel socioeconómico 250-253
- Numerador 61, 262
- Odds
 de la enfermedad 65
 de la exposición 216-217
 ratio 102, 216-219, 238, 259, 328, 336-338, 347
- Ordenadores 437-441
 Introducción de datos 47, 449
- Paquetes para el análisis de la supervivencia 292
- Parámetro (*véase* Variable)
- Percentiles 52
- Pérdida del seguimiento 162, 185, 300, 373
- Periodo de inducción o de latencia 180, 184, 274
- Personas-tiempo a riesgo 66-68, 186, 257
- Placebo 149, 161
- Población 2
 a riesgo 62
 altamente expuesta 175
 diana 145, 227
 de estudio 147, 174
 para un estudio de cohorte 174
 para un estudio de intervención 143
 para un estudio transversal 227
 estándar 77, 87
 experimental 145
 Fracción de exceso 382, 405
 general 174
 Muestras 109, 228
 origen 228
 Registros del cáncer 34, 409-429
 Riesgo atribuible 381
 Riesgo en exceso 382, 405

- Poder estadístico 175, 202, 210, 355-363
- Precisión, cálculos del tamaño de la muestra 363
- Prevalencia 62, 68, 96, 237
- Prevención 9, 377-404
 - primaria 377-387, 423
 - secundaria 377, 387-403, 423
 - terciaria 378, 403, 424
- Probabilidad
 - de morir 282-286
 - de sobrevivir 285
 - Valor de P 117-123, 314, 355
- Protocolo de estudio 32, 432-435
- Prueba
 - de hipótesis 117
 - de Mantel-Haenszel 133, 328-333
 - estadística 133-139, 347-351
 - exacta de Fisher 139
 - de (cuadrado 133-139, 329, 348, 351
 - de tendencia lineal 136-138
- Pseudocáncer 397
- Publicación 451
- Razón
 - de incidencia
 - estandarizada 80, 191
 - proporcional 84, 257
 - de mortalidad
 - estandarizada 80
 - proporcional 84, 257
 - de tasas 102, 217
 - de Mantel-Haenszel 331, 350
 - estandarizada 78
 - de riesgo 102, 217, 329, 349
 - de Mantel-Haenszel 329, 349
- Rechazo 156, 373
- Registros
 - de cáncer 34, 243, 409-429
 - Definición de la población 411-412
 - en países en vías de desarrollo 416-418
 - hospitalarios 34, 409, 427-428
 - poblacionales 34, 409
 - Recopilación de datos 412-414
 - Enlace / Cruzamiento 35, 256
 - ocupacionales 180
- Regresión 267, 340-345
 - de Poisson 343
 - logística 340-345
 - condicional 345
- Repetibilidad, reproducibilidad 26
- Resultado 11, 159-160, 282
- Clasificación errónea 309-310
- Medida 19-20, 184
- Riesgo 64, 68
 - atribuible 103, 378
 - de la población 381
- Diferencia 104
- Exceso 103-105, 378-179, 405
- de la población 382, 405
- Factor 3
- Razón 102, 217, 329, 349
- Reducción 106, 380
- relativo 102
- Seguimiento 164, 185
 - en el registro del cáncer 412, 418, 420
 - en estudios de supervivencia 284
- Pérdida 162, 185, 300, 373
- Sensibilidad 20-26
- Sesgo 149, 295-311
 - de adelanto diagnóstico 396
 - de diagnóstico 221-222
 - de duración 396
 - de medida 161, 184, 194, 301-310
 - de memoria 214, 305
 - de selección 179, 187, 194, 203, 208, 234, 263, 295-300
 - en estudios aleatorizados 300
 - en estudios basados en datos rutinarios 263
 - en estudios caso-control 202, 207-209, 214, 296-299
 - en estudios de cohorte 173, 179-180, 184-185, 194, 299
 - en estudios transversales 234, 296
 - de supervivencia 205
 - del entrevistado 214
 - del observador 214, 306
 - en estudios caso-control 202, 207-208, 214, 296-299
 - en estudios de cohorte 179, 184, 185, 194, 299
 - en estudios de intervención 300
 - en estudios transversales 234, 296
- Significación (*véase* Prueba de Hipótesis)
- Supervivencia
 - Análisis 279-293, 424
 - específica 289
 - Observaciones censuradas 283-284
 - Probabilidad 285
 - Tiempo 283
 - Tendencia 425
 - Sesgo 205
- Sustitutos o próximos 16
- Tablas 56
 - de vida 286.

- Tamaño
 - de muestra 227-237
 - del estudio 353-375
- Tasa
 - acumulativa 82
 - de exposición 14
 - de incidencia 66-68
 - Razón 102, 217
- Técnicas de ciego 161
 - de ciego simple 161
 - de doble ciego 161
 - sin ciego 161
- Tendencias
 - de supervivencia 425
 - según el nivel de duración 189
 - según el nivel de exposición 188
 - temporales 69-75, 253, 261, 425
- Test de (cuadrado 133-139, 329, 348, 351
 - de tendencia lineal 136-138
- Tratamiento 389, 392, 403
- Validez 20-26, 390
- Valor predictivo 21, 390, 394
- Valores extremos 51
- Variable
 - de aproximación 16
 - intermedia 20, 160
- Variación geográfica 243
- Varianza 52
- Voluntarios 145, 296

ISBN: 92-832-0407-7



9 789283 204077