Manual Docente de la Escuela Nacional de Sanidad

Madrid, octubre de 2009

MÉTODO EPIDEMIOLÓGICO

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
Instituto de Salud Carlos III
ENS Escuela Nacional de Sanidad

Escuela Nacional de Sanidad (ENS)
Instituto de Salud Carlos III
Ministerio de Ciencia e Innovación
Sinesio Delgado, 8
28029 MADRID (ESPAÑA)
Tels.: 91 822 22 74
Fax: 91 387 78 56

Catálogo general de publicaciones oficiales:
http://publicaciones.administracion.es

EDITA: ESCUELA NACIONAL DE SANIDAD
Instituto de Salud Carlos III – Ministerio de Ciencia e Innovación

N.I.P.O.: 477–09–019–9
I.S.B.N.: 978–84–95463–56–2

Imprime: Agencia Estatal Boletín Oficial del Estado.
Avda. de Manoteras, 54. 28050 – MADRID

Coordinadores
Miguel Ángel Royo Bordonada
Javier Damián Moreno

Autores
Beatriz Pérez Gómez.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Fernando Rodríguez Artalejo.
Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad Autónoma de Madrid.

Fernando Villar Álvarez.
Departamento de Programas de Salud. Escuela Nacional de Sanidad. Instituto de Salud Carlos III.

Gonzalo López-Abente.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Iñaki Imaz Iglesia.
Agencia de Evaluación de Tecnologías Sanitarias. Instituto de Salud Carlos III.

Javier Damián.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Javier Jiménez Jiménez.
Departamento Médico. AstraZeneca.

Jesús Castilla Catalán.
Servicio de Epidemiología, Prevención y Promoción de la Salud. Instituto de Salud Pública de Navarra.

Jesús González Enríquez.
Agencia de Evaluación de Tecnologías Sanitarias. Instituto de Salud Carlos III.

Jose María Martín Moreno.
Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad de Valencia.

Jose Ramón Banegas Banegas.
Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad Autónoma de Madrid.

Juan de Mata Donado Campos.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Marina Pollán Santamaría.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Miguel Angel Royo Bordonada.
Escuela Nacional de Sanidad. Instituto de Salud Carlos III.

Miguel Delgado Rodríguez.
Departamento de Medicina Preventiva y Salud Pública. Universidad de Jaén.

Nuria Aragonés Sanz.
Centro Nacional de Epidemiología. Instituto de Salud Carlos III.

Para citar este manual

Escuela Nacional de Sanidad (ENS)
Instituto de Salud Carlos III - Ministerio de Ciencia e Innovación.
Miguel Ángel Royo Bordonada, Javier Damián Moreno, *"Método epidemiológico"*. Madrid: ENS - Instituto de Salud Carlos III, Octubre de 2009.

Este texto puede ser reproducido siempre que se cite su procedencia.

ÍNDICE

I. CONCEPTO Y USOS DE LA EPIDEMIOLOGÍA 9
Fernando Rodríguez Artalejo y Jose Ramón Banegas Banegas

II. MEDIDAS DE FRECUENCIA Y DE EFECTO 19
Javier Damián

III. DISEÑO Y TIPOS DE ESTUDIOS EPIDEMIOLÓGICOS 31
Jesús Castilla Catalán

IV. SESGOS Y FACTORES DE CONFUSIÓN 47
Fernando Villar Álvarez

V. ANÁLISIS DE DATOS EPIDEMIOLÓGICOS 75
Javier Damián y Nuria Aragonés

VI. ESTUDIOS DE COHORTES 93
Marina Pollán y Beatriz Pérez

VII. ESTUDIOS DE CASOS Y CONTROLES 117
Javier Jiménez Jiménez

VIII. ESTUDIOS DE PREVALENCIA 125
Juan de Mata Donado Campos

IX. ESTUDIOS ECOLÓGICOS 137
Gonzalo López-Abente

X. ESTUDIOS EXPERIMENTALES 149
Miguel Angel Royo Bordonada y José María Martín Moreno

XI. EPIDEMIOLOGÍA CLÍNICA 169
Iñaki Imaz Iglesia y Jesús González Enríquez

XII. REVISIÓN SISTEMÁTICA Y METAANÁLISIS 187
Miguel Delgado Rodríguez

XIII. INFERENCIA CAUSAL EN EPIDEMIOLOGÍA 207
Jose Ramón Banegas Banegas y Fernando Rodríguez Artalejo

PRÓLOGO

La epidemiología, cuyos primeros antecedentes datan de la época de Hipócrates (siglo IV a.C.) ha ido evolucionando a lo largo del tiempo. Entre los siglos XVII y XX ya se habían realizado de forma aislada algunos estudios de investigación de enfermedades siguiendo un método científico que sentaría las bases del moderno método epidemiológico. La investigación de John Snow, en 1849, sobre la transmisión del cólera por contaminación fecal del agua de consumo, puede considerarse como uno de los más carismáticos de esa época. No obstante, la epidemiología se ha consolidado como una disciplina científica independiente y con identidad propia durante el siglo XX, especialmente a partir del importante despegue en su desarrollo iniciando en la década de los 50. Algunos de los estudios epidemiológicos que han tenido mayor trascendencia pública datan de esa época. Por ejemplo, las investigaciones de Doll y Hill sobre la asociación entre el tabaco y el cáncer de pulmón; los estudios de Keys y Grande Covián sobre la relación entre las grasas de la dieta y los lípidos plasmáticos, que demostraron los efectos perjudiciales de los ácidos grasos saturados y empezaron a sentar las bases de la moderna epidemiología nutricional; los ensayos de intervención comunitaria para evaluar la eficacia de los suplementos de flúor a través de la red de agua de consumo en la prevención de la caries dental; y, por último, el Framingham Study, un estudio de seguimiento diseñado por el Servicio de Salud Pública de los Estados Unidos en la década de los 40 para investigar los factores implicados en la cardiopatía isquémica que, todavía hoy, está proporcionando importantes resultados. La aplicación del método epidemiológico a la investigación de servicios de salud y a la gestión de la práctica clínica es reciente y está llena de oportunidades de desarrollo.

El presente volumen incluye trece temas elaborados por diferentes autores, todos ellos expertos en salud pública y profesores de la Escuela Nacional de Sanidad, con la orientación fundamental de servir de apoyo a la docencia del método epidemiológico. Así, partiendo de la definición de epidemiología y de la descripción de sus usos potenciales, se describen los conceptos básicos de este método científico, sus principales diseños de investigación, los métodos básicos de análisis de datos, las técnicas de control de sesgos y factores de confusión, las bases de la inferencia causal y los aspectos fundamentales de la epidemiología clínica en el escenario actual.

La consulta de diferentes fuentes bibliográficas es especialmente importante en el proceso de aprendizaje del método epidemiológico, una disciplina joven y en pleno desarrollo, con planteamientos y terminologías en ocasiones variables y sujetas a discusión. En este sentido, este manual de apoyo a la docencia incluye, al final de cada capítulo, una selección de lecturas consideradas de referencia, recomendadas como fuentes de consulta para complementar lo expuesto sobre el tema correspondiente.

Miguel Ángel Royo Bordonada y Javier Damián

I. CONCEPTO Y USOS DE LA EPIDEMIOLOGÍA

Fernando Rodríguez Artalejo.
Jose Ramón Banegas Banegas

1. CONCEPTO DE EPIDEMIOLOGÍA

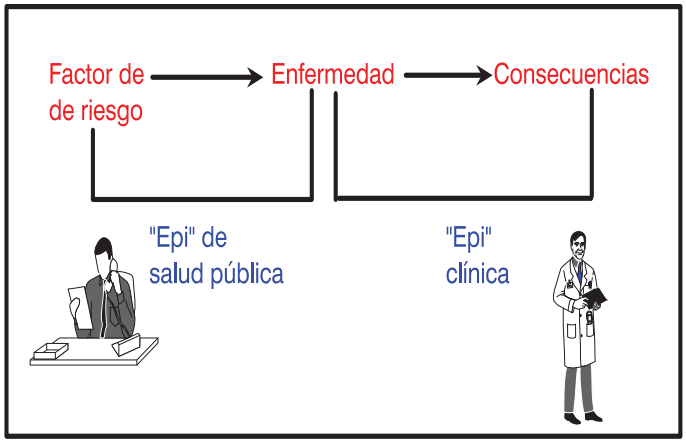
La epidemiología es la disciplina científica que estudia la frecuencia y distribución de fenómenos relacionados con la salud y sus determinantes en poblaciones específicas, y la aplicación de este estudio al control de problemas de salud.

El término *disciplina* alude a un cuerpo de conocimientos que han sido recogidos en libros. El término *científico* se refiere a que dichos conocimientos fueron obtenidos a través de un camino sistemático o método, que pretende garantizar cierta validez y fiabilidad. El *estudio* incluye las investigaciones caracterizadas por la simple vigilancia y observación de fenómenos para medir su magnitud y sugerir hipótesis sobre su origen. Este tipo de investigaciones reciben el calificativo de *descriptivas*. Pero también incluye las investigaciones dirigidas a contrastar estas hipótesis mediante estudios observacionales y experimentales. Estas investigaciones reciben el nombre de *analíticas*. *Distribución* significa la medida de la frecuencia y variación de un fenómeno en grupos de población a lo largo del *tiempo*, en diferentes *lugares* o formados por diferentes tipos de *personas*.

La epidemiología no sólo estudia enfermedades sino *todo tipo de fenómenos relacionados con la salud*, entre los que se encuentran causas de muerte como los accidentes o suicidios, hábitos de vida como el consumo de tabaco o la dieta y el uso de servicios de salud o la calidad de vida relacionada con la salud, entre otros. Los *determinantes* de estos fenómenos son todos los factores físicos, biológicos, sociales, culturales y de comportamiento que influyen sobre la salud. Los fenómenos relacionados con la salud y sus posibles determinantes dan lugar a algunas de las clasificaciones de las ramas de la epidemiología. Así, cuando el eje de clasificación son los fenómenos sanitarios surgen ramas como la epidemiología cardiovascular, del cáncer, o de los servicios sanitarios. Cuando el eje son los determinantes, surgen la epidemiología nutricional, laboral, o social.

Por último, la epidemiología es una disciplina básica de la salud pública y de la medicina clínica, porque sus conocimientos pueden y deben ser aplicados al control de problemas de salud en ambos campos. Estas aplicaciones de la epidemiología permiten su clasificación en dos ramas. La epidemiología general o de salud pública y la epidemiología clínica. La distinción entre ambas ramas no estriba tanto en las técnicas utilizadas como en la porción de la historia natural de la enfermedad que es estudiada por cada una de ellas. La historia natural de una enfermedad es el conjunto de sucesos que van desde que un sujeto o grupo de sujetos resulta expuesto a las primeras causas de una enfermedad hasta que ésta se desarrolla y finalmente se resuelve con la curación total, la curación con secuelas o la muerte (figura 1-1). La *epidemiología de salud pública* estudia la primera parte de esta cadena de sucesos, es decir, la frecuencia y distribución de la enfermedad y sus determinantes, factores de riesgo o protección. Para ello se fija en sujetos sanos, generalmente viviendo en la comunidad, a los que sigue para observar como enferman. La *epidemiología clínica* estudia la frecuencia y distribución de las consecuencias de la enfermedad y sus determinantes, los factores pronósticos. Para ello, suele fijarse en sujetos enfermos en los que mide posibles factores pronósticos y los sigue para observar la evolución de la enfermedad.

Figura 1-1. Representación esquemática de las relaciones habituales de estudio por la epidemiología de salud pública y la epidemiología clínica a lo largo de la historia natural de la enfermedad.



Sin embargo el rasgo que más diferencia a la epidemiología de otras disciplinas biológicas es el estudio de la frecuencia de fenómenos en grupos de sujetos, en *poblaciones*. Mientras que las ciencias experimentales en el laboratorio estudian, sobre todo, relaciones *deterministas* o no estocásticas, la epidemiología se centra en relaciones *probabilísticas* o estocásticas. En epidemiología, los potenciales factores etiológicos pueden, en un caso concreto, producir o no producir una enfermedad. Así, una persona que fuma puede o no desarrollar un infarto agudo de miocardio, siendo la segunda posibilidad algo más frecuente que la primera. Nuestros factores causales aumentan la frecuencia o probabilidad de desarrollar una enfermedad -son factores de riesgo- pero no aseguran su desarrollo. Para poner de manifiesto su efecto se necesitan grupos de personas en los que medir la frecuencia de la enfermedad y compararla entre los expuestos y no expuestos al factor etiológico. En los estudios experimentales de laboratorio se suelen abordar relaciones deterministas. En ellas el factor causal es suficiente para producir su efecto. Siempre que alguien se exponga a él sufrirá su efecto. Para ponerlo de manifiesto bastará con estudiar un sólo sujeto.

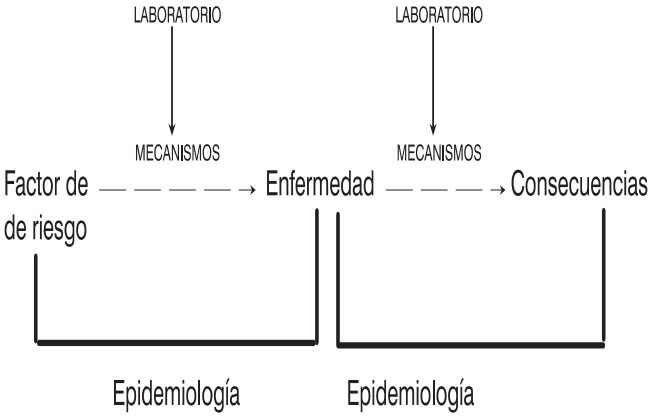
La distinción entre relaciones probabilistas y deterministas es conocida desde antiguo. Un ejemplo ilustrativo, que es utilizado con frecuencia para describir este aspecto, se encuentra en el estudio de los efectos de las radiaciones ionizantes, uno de los factores etiológicos más investigados por la ciencia, tanto en el laboratorio como fuera del mismo. Dosis altas con elevadas tasas de exposición producen efectos deterministas. Dosis muy bajas recibidas de forma crónica dan lugar a efectos probabilísticos. Las características de estos tipos de efectos se presentan en la tabla 1-1. Esta tabla también sugiere que los efectos deterministas suelen tener tiempos de inducción cortos. Ello permite que un experimento para medirlos requiera sólo minutos, horas o algunos días, a diferencia de lo que ocurre en epidemiología, donde los estudios pueden durar años.

Tabla 1-1. Resumen de las principales diferencias entre los efectos estocásticos y los no estocásticos

EFECTOS ESTOCÁSTICOS	EFECTOS NO ESTOCÁSTICOS
- Probabilidad del efecto depende de la dosis de exposición.	- Probabilidad del efecto NO depende de la dosis de exposición. La probabilidad es cero o uno.
- Intensidad del efecto NO depende de la dosis de exposición.	- Intensidad del efecto depende de la dosis de exposición.
- Efecto SIN umbral de dosis de exposición.	- Efecto CON umbral de de dosis de exposición.
- Frecuentemente tiempo de inducción LARGO.	- Frecuentemente tiempo de inducción CORTO.

Típicamente los científicos de laboratorio exploran los numerosos mecanismos patogénicos o de resistencia de la enfermedad. Los mecanismos, a modo de los componentes de un engranaje, tienen un carácter inmediato o cercano entre sí y, por tanto, las relaciones de interés pueden verse como deterministas. Por su parte, el epidemiólogo estudia de una sola "tacada" la relación entre variables más o menos distales y la aparición de enfermedad y su evolución (figura 1-2). Las variables son tan lejanas entre sí que sus efectos tardan más tiempo en producirse -tiempo de inducción largo- y sólo se concretan en presencia de terceras variables causales e intermedias. Por ello, las relaciones entre dichas variables son probabilistas. Cuanto más se alejan las variables de resultado de las de exposición, las relaciones se hacen además menos específicas. Un ejemplo de ello se encuentra en la epidemiología social, donde los factores de exposición, habitualmente variables socioeconómicas, tiene un efecto variable y con tiempos de inducción también diferentes sobre un amplio espectro de enfermedades. Junto a estos abordajes extremos, también observamos otros intermedios y complementarios, en los que las variables de exposición en epidemiología se acercan, en la cadena causal, a las variables resultado. Es el caso de la epidemiología molecular. Por último los dos abordajes de estudio son complementarios. La mayoría de los agentes nocivos para la salud que hoy conocemos y que son modificables han sido identificados por estudios epidemiológicos. Es el caso del tabaco, la hipertensión arterial, la hipercolesterolemia, etc.. La epidemiología representa a menudo la única posibilidad de estudiar directamente en humanos los efectos sobre la salud de potenciales factores etiológicos, evitando la extrapolación de resultados desde animales y otros modelos de investigación al hombre. Sin embargo, estos estudios enriquecen sus resultados con la plausibilidad biológica derivada de los abordajes de laboratorio.

Figura 1-2. Representación esquemática de las relaciones habituales de estudio por la epidemiología y las disciplinas de laboratorio a lo largo de la historia natural de la enfermedad.



2. USOS DE LA EPIDEMIOLOGÍA EN EL CONTROL DE LOS PROBLEMAS DE SALUD

El fin último de la epidemiología es controlar los problemas de salud. Para ello la epidemiología genera información en tres campos.

Tabla 1-2. Algunos tipos de información que se deriva de estudios epidemiológicos

CONOCIMIENTO DE LA HISTORIA NATURAL DE LAS ENFERMEDADES O PROBLEMAS DE SALUD

- Sugerir hipótesis sobre los posibles factores de riesgo de una enfermedad.
- Identificar y medir el efecto de dichos factores sobre la aparición de la enfermedad.
- Medir el tiempo de inducción de la acción de algún factor de riesgo.
- Medir la evolución (supervivencia, tiempo hasta la desaparición de discapacidad o de secuelas) de una enfermedad.
- Identificar y medir el efecto de dichos factores pronósticos.
- Evaluar la eficacia y efectividad de un abordaje diagnóstico o terapéutico de una enfermedad.

CONOCIMIENTOS PARA EL CONTROL DE PROBLEMAS DE SALUD EN LAS POBLACIONES, INFORMANDO Y EVALUANDO LOS DISTINTOS PASOS DE LO QUE GENERICAMENTE SE CONOCE COMO PLANIFICACION SANITARIA

- Medir la importancia sociosanitaria de un problema de salud, con alusión especial a la exposición a factores de riesgo o sus consecuencias de morbilidad, discapacidad o muerte.
- Formular objetivos de salud cuantificables y con una referencia temporal.
- Medir recursos sanitarios disponibles, su utilización y los recursos necesarios para satisfacer necesidades de salud.
- Establecer prioridades de actuación.
- Establecer el beneficio esperado de una intervención sanitaria.
- Evaluar el progreso en el control de una enfermedad a nivel comunitario.
- Evaluar un programa de salud o intervención sanitaria.
- Validar un sistema de información sanitaria.

CONOCIMIENTOS PARA EL CONTROL DE PROBLEMAS DE SALUD EN LOS INDIVIDUOS, MEJORANDO EL MANEJO CLINICO DE LOS PACIENTES ENFERMOS O DE LOS SANOS DE ALTO RIESGO

- Establecer la idoneidad de una prueba diagnóstica y el orden en que debe aplicarse en el contexto de otras pruebas.
- Establecer la prueba diagnóstica o el tratamiento más útil para un sujeto concreto.
- Elaborar protocolos clínicos.

Primero, sobre la historia natural de las enfermedades y la eficacia de medidas preventivas y curativas que pretenden modificar dicha historia de forma más favorable para el ser humano. Segundo, para formular, ejecutar y evaluar planes y programas de salud que mejoren el nivel de salud de las poblaciones. Tercero, para mejorar el proceso de toma de decisiones clínicas, dirigidas a mejorar la salud de sujetos enfermos y al desarrollo de protocolos o guías clínicas (tabla 1-2).

2.1. Conocimiento de la historia natural de la enfermedad

La epidemiología permite conocer muchos elementos de la historia de la enfermedad. Sólo destacaremos algunos de los más importantes. La epidemiología identifica la/s *causa/s de la enfermedades*. El estudio Framingham permitió conocer que el tabaquismo, la presión arterial, la colesterolemia, la intolerancia a la glucosa o la hipertrofia ventricular izquierda son algunos de los principales factores de riesgo de la cardiopatía isquémica. Para ello, se siguió a un grupo de personas inicialmente libres de la enfermedad y se observó que a mayor valor de las anteriores variables más alta era la frecuencia con que se desarrollaba una enfermedad coronaria a lo largo del seguimiento.

La epidemiología puede medir el *tiempo de inducción* o incubación de la enfermedades. En las enfermedades infecciosas basta calcular la diferencia entre el momento en que aparece la enfermedad y aquél en el que se produjo el contagio por el agente de la infección. El tiempo de incubación es útil, entre otros propósitos, para establecer la cuarentena o tiempo durante el que las personas potencialmente infectadas por un agente de la infección deben mantenerse fuera del contacto con el resto de la comunidad para evitar el contagio. La cuarentena ha de durar como mínimo el tiempo de inducción de la enfermedad. Si en ese periodo la persona no ha desarrollado la enfermedad puede ya convivir de forma más cercana con el resto de la comunidad sin peligro de transmisión de la enfermedad, pues realmente no está infectado. En el caso de las enfermedades crónicas no infecciosas, en las que la exposición al factor de riesgo dura muchos años, la medida del tiempo de inducción es algo más difícil. Rose observó que la correlación de la colesterolemia con la mortalidad por cardiopatía isquémica en las cohortes del estudio de siete países, aumentaba progresivamente de los 5 a los 10 años, y de éstos a los 15 años de seguimiento. Concluyó que el tiempo de inducción del efecto de la colesterolemia sobre la mortalidad coronaria es al menos de 15 años.

Medir el *pronóstico* de una enfermedad es relativamente fácil. Basta con seguir a lo largo del tiempo a un grupo de sujetos para estimar su supervivencia. Así, Marrugat y cols. siguieron una cohorte de enfermos coronarios y registraron el número de muertos que se producían en la misma a lo largo del tiempo. Ello, les permitió establecer que, en nuestro medio, la supervivencia del infarto agudo de miocardio a los 10 años era aproximadamente el 55%. La creación de registros, por ejemplo de enfermedades cardiovasculares, facilita la información periódica sobre la supervivencia de las mismas y su mejora a lo largo de los años, como consecuencia principalmente de los avances tecnológicos y la mejor asistencia sanitaria.

La epidemiología también permite medir la eficacia de medidas preventivas y terapéuticas, es decir, sus beneficios cuando se administran en las mejores condiciones posibles de trabajo en el sistema sanitario, con pacientes que reúnen en grado sumo las indicaciones de tratamiento y un grado de cumplimiento terapéutico óptimo. Así un ensayo clínico, un diseño epidemiológico de tipo experimental, ha permitido comprobar que la administración de simvastatina, un fármaco hipocolesterolemiante, a sujetos con elevaciones moderadas de la colesterolemia y con enfermedad coronaria previa, reduce la frecuencia de infartos de miocardio y la mortalidad general. En concreto, el tratamiento con simvastatina durante 5 años reduce aproximadamente un 30% la mortalidad general. También es posible estimar la *efectividad* de las intervenciones sanitarias, es decir, sus beneficios cuando se administran en las condiciones usuales de práctica

clínica. Esta actividad de investigación recibe el nombre de *"outcomes research"* o investigación sobre las consecuencias de la práctica clínica. Por este tipo de investigación sabemos que el grado de conocimiento, tratamiento y control de la presión arterial en España es todavía muy mejorable, que hay una gran variabilidad en la mortalidad intrahospitalaria por cardiopatía isquémica en España, o que la calidad de vida de los pacientes tratados mediante derivación aortocoronaria mejora sustancialmente al año de la intervención, a pesar de que el control de los factores de riesgo de estos pacientes es subóptimo. También este tipo de investigación ha mostrado que algunos avances terapéuticos (antiagregantes plaquetarios, betabloqueantes, etc..) han reducido la letalidad del infarto agudo de miocardio en España en el periodo comprendido entre 1978 y 1993.

2.2. Control de problemas de salud en las poblaciones

La epidemiología proporciona información para responder a las tres preguntas *"hipocráticas"* de la prevención.

2.2.1. La enfermedad o problema de salud, ¿es prevenible o controlable?

La epidemiología ha demostrado que muchas enfermedades son prevenibles o controlables. Asimismo, ha identificado los determinantes de los problemas de salud y ha comprobado que la intervención es posible y además eficaz. Por todo ello, es considerada una ciencia básica de la medicina preventiva y la salud pública. La epidemiología cardiovascular ha experimentado un gran desarrollo tras la Segunda Guerra Mundial y constituye un buen ejemplo para ilustrar muchas de las contribuciones, y también alguna limitación, de esta ciencia para el control de las enfermedades.

Las evidencias sobre la "controlabilidad" de las enfermedades cardiovasculares proceden de estudios sobre las variaciones geográficas de la mortalidad por esta enfermedad, sus cambios en una misma comunidad a lo largo del tiempo y los estudios en emigrantes, en los que éstos adoptan las tasas de mortalidad cardiovascular (y presumiblemente también los hábitos de vida) de las comunidades que los acogen. En ausencia de cambios en la composición genética de los sujetos, estos estudios atribuyen la variación en la mortalidad a factores ambientales sobre los que es posible intervenir. Asimismo informan sobre la magnitud potencial del descenso de la mortalidad tras la intervención y el tiempo máximo necesario para obtener este beneficio. La gran variación en la distribución provincial de la mortalidad cardioisquémica y cerebrovascular en España sugiere que es teóricamente posible reducir a cerca de la mitad la mortalidad en las provincias que la tienen más elevada. Un paso más allá lo da un estudio reciente que muestra que la ingesta de folatos puede estar estar relacionada con la distribución geográfica de la enfermedad coronaria en España. Un argumento similar al construido para España puede hacerse a nivel europeo, donde las variaciones geográficas en la mortalidad cardiovascular tienen una gran magnitud; la mortalidad cardiovascular de algunas regiones europeas es 4 ó 5 veces superior a la de las regiones con menor mortalidad. Continuando con las evidencias de la controlabilidad de la enfermedad cardiovascular, el descenso de la mortalidad por cardiopatía isquémica en Canadá en las últimas cuatro décadas demuestra que ha sido posible reducir la mortalidad por estas enfermedades entre los varones en un 50% en un plazo tan corto como 20 años. Evidencias similares a esta última existen para Japón, Finlandia, Estados Unidos de América y otros países desarrollados.

2.2.2. Si la enfermedad es prevenible, ¿Cuales son las estrategias de prevención y control más adecuadas?

Esta pregunta se ha respondido en el caso de las enfermedades cardiovasculares a través de dos grandes tipos de estudios: los observacionales y los experimentales. Históricamente apareció

primero una cohorte de estudios observacionales que, desde el final de los años 40 hasta la mitad de los 60, generaron grandes avances en la identificación de los llamados factores de riesgo mayores de estas enfermedades. De hecho el término, y probablemente también el concepto, de factor de riesgo fue utilizado por vez primera por William Kannel en 1961 en una de las publicaciones del estudio Framingham. De forma similar, fue Ancel Keys el que primero sugirió que los "estilos de vida" eran factores de riesgo de las enfermedades crónicas, en concreto las cardiovasculares, con motivo de las investigaciones que dieron posteriormente lugar al estudio de "los siete países".

A partir de la década de los 60 se realizaron estudios experimentales de la eficacia de diversas intervenciones sobre los hábitos de vida para el control de las enfermedades cardiovasculares. Estos estudios tienen más validez que los observacionales y además pueden proporcionar evidencia directa sobre dos aspectos relevantes para la prevención de las enfermedades: primero, demuestran que la intervención es factible, segundo, demuestran que es eficaz.

Los estudios anteriores, realizados tanto sobre individuos como poblaciones, crearon las condiciones para que Geoffrey Rose propusiera dos estrategias polares de prevención de enfermedades. La primera de ellas consiste en identificar a sujetos en alto riesgo de enfermar, para reducir en ellos la exposición a los factores de riesgo. La segunda actúa sobre toda la población de interés y pretende trasladar hacia la izquierda el conjunto de la distribución de los factores de riesgo. Esta estrategia tiene sentido porque los primeros estudios observacionales demostraron que los principales factores de riesgo tienen una relación gradual y continua con el riesgo cardiovascular. Asimismo, y porque se ha demostrado que los distintos factores de riesgo cardiovascular interaccionan para producir su efecto, parece conveniente que la reducción del riesgo cardiovascular, que es el objetivo principal de las estrategias de prevención, requiera la actuación simultánea sobre todos los factores. Por último, y aunque ambas estrategias de riesgo no son excluyentes sino complementarias, es la poblacional la que presumiblemente reportará más beneficios. Ello se debe a que la mayor parte de los casos de enfermedad no ocurren entre los relativamente pocos sujetos de alto riesgo de una población, sino entre los mucho más numerosos que tienen unos valores de los factores de riesgo que, a efectos didácticos, podríamos calificar de "normales o ligeramente elevados".

2.2.3. ¿Cual es la magnitud del beneficio de las estrategias de prevención?

Es quizás esta información la más importante para conseguir que una estrategia de prevención pueda llevarse a la práctica una vez formulada. A partir de información de la literatura internacional y de datos de prevalencia de los principales hábitos de vida en España, Banegas y cols. calcularon la proporción de la mortalidad cardiovascular en nuestro país que es atribuible a la exposición simultánea a los principales factores de riesgo conocidos. De acuerdo con sus resultados, cerca del 50% de la mortalidad cardiovascular es potencialmente prevenible en España, si se controla la colesterolemia, la presión arterial, el hábito tabáquico, la obesidad y el sedentarismo. Este mismo grupo ha calculado el beneficio potencial del control de la colesterolemia en España realizado de forma realista a través de estrategias individual y de población. A corto plazo la estrategia poblacional podría reducir la mortalidad cardioisquémica en un 10% frente a un 3,3% de reducción debida a la estrategia individual. La reducción de esta mortalidad potencialmente debida a la aplicación conjunta de ambas estrategias sería de un 12,5%.

Por último, otra pieza de información importante es la distribución de los beneficios de prevención a lo largo del tiempo. Así, se ha simulado la reducción de la mortalidad por diversas enfermedades que es posible obtener año a año, desde el presente hasta el 2020, a través del control del hábito tabáquico en España. El resultado más llamativo de este trabajo es que una

reducción del 40% del hábito tabáquico a lo largo del periodo 1992 a 2000 se traduciría en una ganancia de aproximadamente 60.000 años de vida en el año 2020.

Una vez se ha respondido afirmativamente a estas tres cuestiones, se está en condiciones de llevar a cabo una estrategia o plan para el control de la enfermedad o problema de salud en la población. Un plan consiste en la formulación de unos objetivos de salud y las acciones para conseguirlos. La epidemiología proporciona información para la elaboración de un plan. La definición de objetivos se produce una vez medidas las necesidades de salud de la población y establecido cuáles de ellas se deben atender de forma prioritaria. La medida de necesidades se realiza por procedimientos epidemiológicos, como la obtención de indicadores de salud (sociodemográficos, sanitarios, de utilización de servicios de salud o de recursos sanitarios disponibles) y de encuestas de salud, que miden la morbilidad percibida por la población, sus hábitos de vida con relevancia para la salud, y el uso de servicios de salud. Se dispone además de indicadores como los años de vida potencialmente perdidos, que miden el impacto de un determinado problema sobre la salud de las poblaciones.

Las prioridades se establecen en función de la intensidad y extensión de las necesidades y de los medios disponibles para satisfacerlas. La magnitud de un problema de salud medido por la incidencia de una enfermedad y su evolución a lo largo del tiempo, su gravedad medida por la letalidad o las discapacidades que produce, y su extensión medida por la distribución en grupos de personas o de lugares, son elementos que junto a la sensibilidad del problema de salud a la intervención sanitaria, los costes y factibilidad de la misma, los valores de las personas que toman las decisiones y la coyuntura social y económica contribuyen a la priorización de un problema.

Hay dos formas principales por las que la epidemiología es útil para formular objetivos cuantificados y con un marco temporal. Primero, proyectando la tendencia de un indicador en el futuro para establecer cuál será el valor probable del mismo en ausencia de intervención. Segundo, modificando el valor del mismo en función de la eficacia que han demostrado las intervenciones sanitarias sobre ese valor.

La epidemiología también es útil para evaluar las acciones desarrolladas para conseguir los objetivos de salud. En primer lugar, permite establecer por medio de estudios de vigilancia y monitorización en qué medida las acciones se han ejecutado de la forma en que estaba previsto (evaluación de proceso). En segundo lugar, por estudios descriptivos y fundamentalmente analíticos valora si los objetivos formulados han sido alcanzados (evaluación de resultados). Un ejemplo de evaluación llevada a cabo por técnicas descriptivas es el progreso realizado en la consecución de los 38 objetivos de la estrategia de Salud para Todos, uno de los cuales corresponde a las enfermedades cardiovasculares. Un ejemplo de evaluación realizada por técnicas analíticas a nivel individual es la determinación de la factibilidad y eficiencia del programa de cribado de familiares de casos índice de hipercolesterolemia familiar o la calidad de la asistencia a los pacientes coronarios. Otro ejemplo de evaluación, esta vez a nivel comunitario (estudio cuasi-experimental), es la observación de una reducción en la mortalidad por enfermedad cardiovascular (y de cáncer), asociada a una disminución de la prevalencia de sus factores de riesgo biológicos (dislipemia, hipertensión) y de una mejora en los hábitos de vida (reducción del tabaquismo y de la ingesta de grasas saturadas) en la región de Karelia del norte (sometida a intervención de control de riesgo cardiovascular) por comparación con lo ocurrido en la región de Kuopio (no sometida a intervención). La evaluación de programas de reducción de riesgo cardiovascular adquirirá gran auge en los próximos años, al generalizarse los programas integrados de atención cardiovascular, a nivel de organizaciones, regiones o países enteros.

Por último, relacionado con la formulación y ejecución de programas está el establecimiento de estándares de calidad de la atención cardiovascular y la monitorización, mediante técnicas epidemiológicas, del progreso en la consecución de dichos estándares. También es muy importante la validación de los sistemas de información sanitaria que son útiles para la evaluación.

2.3. Control de problemas de salud en los individuos

Quizás el objetivo más importante de la actividad asistencial es mejorar el curso clínico de los enfermos. Ello se consigue con un diagnóstico y tratamiento correctos. Sin embargo, son varias las posibilidades que se le ofrecen al médico para cada paciente. Estas posibilidades incluyen utilizar o no algún procedimiento diagnóstico o terapéutico y, de hacerlo, elegir el más adecuado entre varios posibles. Del acierto de estas decisiones dependerá en buena medida restablecer la salud del enfermo o mejorar el pronóstico del sujeto sano con alto riesgo de enfermar.

Se dispone de un grupo de técnicas, como las basadas en los umbrales de probabilidad o los árboles de decisión, agrupadas genéricamente bajo el nombre de técnicas de análisis de decisión, que pueden ayudar al médico clínico a mejorar su trabajo en beneficio de sus pacientes. A partir de ellas, y complementadas con las revisiones sistemáticas de la literatura (meta-análisis), se pueden elaborar los llamados protocolos o guías clínicas, que son un conjunto de recomendaciones para el manejo clínico de grupos de pacientes similares. Probablemente, pocas áreas de la medicina están mejor preparadas que la medicina cardiovascular para el ejercicio de la Medicina Basada en la Evidencia, debido al gran desarrollo de la investigación en este campo en los últimos años, que ha dado lugar a evidencias de alta calidad.

3. REFERENCIAS BIBLIOGRÁFICAS

GÓMEZ LÓPEZ LI, RABANAQUE HERNÁNDEZ MJ, AIBAR REMÓN C. *Diseño de programas de salud.* En: Medicina Preventiva y Salud Pública. Barcelona: Masson, 2001: 1047-1062.

LAST JM, ED. *A dictionary of epidemiology.* 4ª ed. Nueva York: Oxford University Press, 2000.

MARRUGAT J, SALA J. *Registros de morbimortalidad en cardiología: metodología.* Rev Esp Cardiol 1997; 50: 48-57.

PINEAULT R, DAVELUY J. *La planificación sanitaria: conceptos, métodos, estrategias.* Barcelona: Masson-SG, 1990: 43-211.

REGIDOR E, GUTIÉRREZ-FISAC JL. *Indicadores de Salud. Cuarta evaluación en España del programa regional europeo Salud para Todos.* Madrid: Ministerio de Sanidad y Consumo, 2000.

REY CALERO J. *Epidemiología y epidemiología clínica.* En: Rey Calero J, Herruzo Cabrera R, Rodríguez Artalejo F. Fundamentos de epidemiología clínica. Madrid: Síntesis, 1996: 25-35.

RODRÍGUEZ ARTALEJO F, FIGAREDO ALVARGONZÁLEZ C, HERNÁNDEZ VECINO R. *Investigación de resultados.* En: del Llano Señaris J, Ortún Rubio V, Martín Moreno JM, Millán Núñez-Cortés J, Gené Badía J, eds. Gestión Sanitaria: Innovaciones y desafíos. Barcelona: Masson-MSD, 1997:529-542.

RODRÍGUEZ ARTALEJO F, GUTIÉRREZ-FISAC JL. *El estado de salud y sus determinantes. Evaluación de los objetivos 1 a 12.* En: Alvarez Dardet C, Peiró S, eds. La salud pública ante los desafíos de un nuevo siglo. Informe SESPAS 2000. Granada: Escuela Andaluza de Salud Pública, 2000: 43-50.

RODRÍGUEZ ARTALEJO F, ORTÚN RUBIO V, BANEGAS JR, MARTÍN MORENO JM. *La epidemiología como instrumento para una política de salud racional.* Med Clin (Barc) 1989; 93: 663-666.

ROSE G. *Incubation periods of coronary heart disease.* BMJ 1982; 284: 1600-1601.

ROSE G. *La estrategia de la medicina preventiva.* Barcelona: Salvat-Masson, 1994.

II. MEDIDAS DE FRECUENCIA Y DE EFECTO

Javier Damián

1. MEDIDAS DE FRECUENCIA DE ENFERMEDAD

Existen dos grupos de medidas según la consideración del tiempo: medidas de incidencia y de prevalencia. Las medidas de *incidencia* de enfermedad tienen en cuenta los casos nuevos que se producen en poblaciones. Las medidas de *prevalencia* consideran los casos de enfermedad existentes en una población en un momento o periodo.

1.1. Incidencia

Existen dos tipos importantes de medidas de incidencia: la incidencia acumulada (o proporción de incidencia) y la tasa de incidencia. La *incidencia acumulada (IA)* relaciona el número de casos con el tamaño poblacional al comienzo de un periodo de tiempo (Δt):

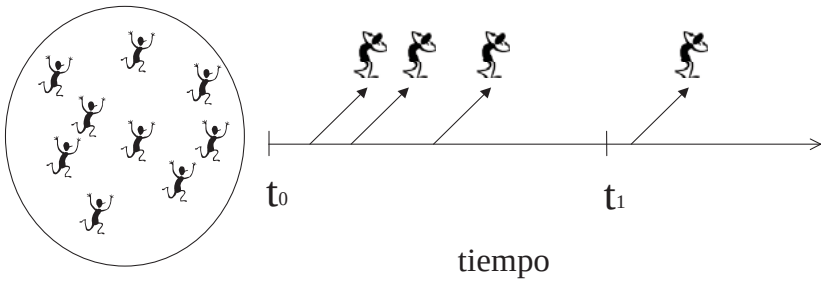
$$IA = \frac{\text{nº de casos nuevos en } \Delta t}{\text{nº de personas al inicio del periodo } (t_0)}.$$

Esta medida por tanto supone la *proporción* de personas que han enfermado en ese periodo.

En la Figura 2-1 se observa una población de 9 personas al inicio del seguimiento. Durante el seguimiento se van produciendo ocurrencias de enfermedad. La incidencia acumulada para un periodo de t_1 es de 3 casos nuevos /9 personas al inicio del seguimiento = 0,33.

Como toda *proporción* no tiene unidades y sus posibles valores están entre 0 y 1. Otra característica de las proporciones es que pueden interpretarse en términos de *probabilidad*: la probabilidad de desarrollar una enfermedad en un periodo de tiempo. También se usa el término *riesgo*. Pero para que pueda estimarse el riesgo es necesario que las personas del grupo al inicio de seguimiento *puedan* desarrollar la enfermedad (sean susceptibles). Si quisiéramos medir la incidencia acumulada de sarampión en las 9 personas del ejemplo anterior y una tuviera inmunidad deberíamos excluirla del denominador, siendo la incidencia acumulada de 3/8 = 0,37.

Figura 2-1. Seguimiento de una cohorte de 9 personas.



Otra asunción que hay que hacer para poder estimar adecuadamente el riesgo es que no haya muertes por otra causa que no sea el evento que se está midiendo (riesgos competitivos), ya que si esto ocurre no se sabrá si las personas hubieran desarrollado la enfermedad caso de no haber muerto. En general se tiene que cubrir el periodo en riesgo. Por tanto la incidencia acumulada puede entenderse como una estimación de la probabilidad *condicional* de desarrollar el evento en un tiempo determinado.

Para interpretar adecuadamente un riesgo es necesario mencionar el tiempo de seguimiento ya que el valor de la incidencia acumulada tiende a aumentar (tiende a 1) conforme este tiempo es mayor. Si pasa tiempo suficiente la probabilidad de ocurrencia de cualquier evento aumenta. No tiene sentido informar de una incidencia acumulada de, por ejemplo, 12% ya que puede ser muy grande si se refiere a un mes o pequeña si se refiere a 10 años.

La incidencia acumulada se mide sobre una *cohorte* que es, estrictamente hablando, una *población fija o cerrada*. Este tipo de poblaciones se caracterizan por no permitir entradas de nuevos miembros, sólo salidas: porque mueren, porque desarrollan la enfermedad, y por otros motivos.

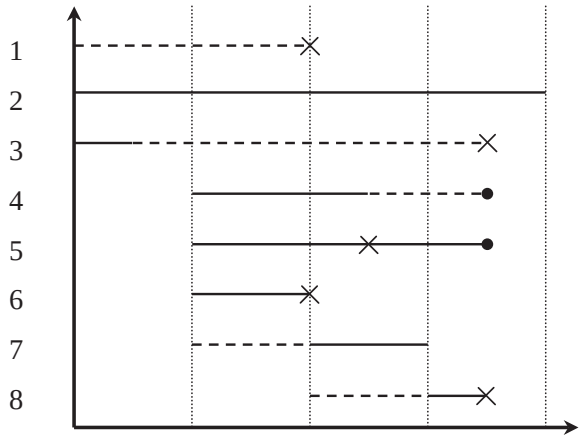
Otro problema es que en los estudios es habitual que haya sujetos que se incorporen más tarde por lo que los tiempos de seguimiento son diferentes.

La *tasa de incidencia, I*, basada en denominadores de *persona-tiempo* permite analizar datos obtenidos del seguimiento de *poblaciones dinámicas* (permiten entradas y salidas). La tasa de incidencia se define como el número de casos nuevos por cada unidad persona-tiempo de observación:

$$I = \frac{\text{nº de casos nuevos}}{\text{nº de personas-tiempo en observación}}.$$

El número de personas-tiempo se obtiene sumando los tiempos en observación que cada sujeto de estudio está *en riesgo* de ocurrencia del evento que se trata de medir. Nótese que el denominador de la tasa de incidencia es bidimensional, compuesto no sólo por el tamaño poblacional sino también por la cantidad de tiempo de observación. La inclusión de la dimensión temporal nos ofrece una medida de la 'velocidad' con que ocurre una enfermedad en una población.

Figura 2-2. Población dinámica de 8 personas a lo largo de 20 años.
La línea discontinua representa consumo moderado de alcohol. La continua consumo cero.
Las aspas representan casos nuevos de infarto de miocardio. Los puntos negros son muertes.



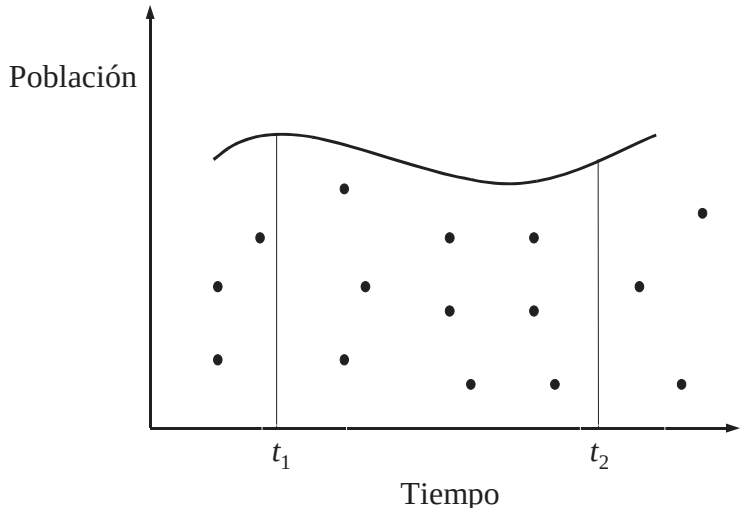
En la Figura 2-2, se representa el seguimiento durante 20 años a una hipotética población dinámica de 8 personas para medir tasas de incidencia de infarto de miocardio, representado por un aspa. La cantidad persona-tiempo de observación que aporta la persona nº 1 es de 10 años; la persona nº 2 está en riesgo 20 años. La 3, 17,5 años; la 4, 12,5 años. La persona nº 5 aporta 7,5 años en riesgo; la 6, 5 años; la 7, 10 años; la persona nº 8 aporta 7,5 años. Todo suma 90 años de observación en riesgo, por lo que la tasa de incidencia es de 5/90 = 0,056 año^{-1} o 5,6 casos por cada 100 personas-año.

Al contrario que la incidencia acumulada la tasa de incidencia sí tiene unidades: el inverso del tiempo (*1/tiempo, o tiempo*$^{-1}$). Consecuentemente el inverso de la tasa se interpreta como el tiempo medio hasta la ocurrencia de enfermedad. Además tiene un *rango* de valores entre 0 e infinito (no tiene un límite superior teórico).

La tasa de incidencia también ha sido denominada como *densidad de incidencia*. Esta idea se aprecia bien en la Figura 2-3, tomada de Morgenstern y cols., que representa los casos que se producen en un 'mar población-tiempo'. Es fácil apreciar la idea de densidad y también que en una población estable se pueden medir la tasas de incidencia, es decir el número de casos nuevos por unidad de área, o cantidad persona-tiempo, con información demográfica secundaria (censos, movimiento natural de la población, registros...).

Lo importante en el cómputo de personas-tiempo es que, al igual que en el caso de la incidencia acumulada, sea *en riesgo* de ocurrencia de casos nuevos. Además debe representar adecuadamente a la población de interés. A diferencia de la población cerrada una misma persona puede aportar tiempo en riesgo a, o ser parte de, más de una población diferente, por ejemplo cambiando de grupo de edad o de exposición (pasar a ser ex-fumador). En el ejemplo anterior (Figura 2-2) la parte discontinua representa consumo moderado de alcohol y la parte continua consumo cero. Hay personas que aportan tiempo en riesgo a ambas poblaciones (3, 4, 7 y 8).

Figura 2-3. Población dinámica seguida en el tiempo. Los puntos son casos nuevos de eventos de interés.

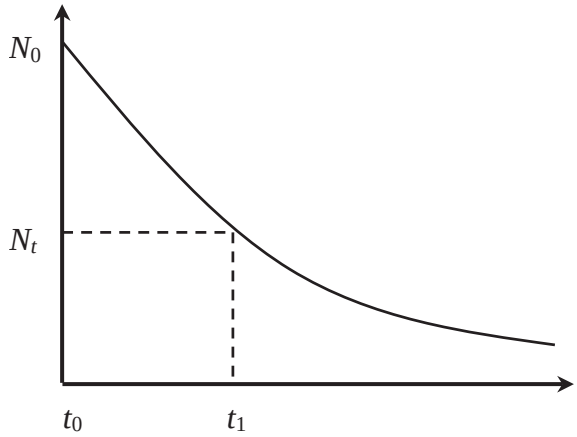


Normalmente el interés fundamental es la medición de primeros episodios aunque no siempre. Ejemplos como caídas o resfriados pueden requerir la contabilización de más de un episodio por persona aunque los métodos en estos casos suelen requerir mayor complejidad analítica.

1.2. Relaciones entre tasa de incidencia e incidencia acumulada

La incidencia acumulada se mide sobre una cohorte fija. En la Figura 2-4 se representa el tamaño de una cohorte fija. Con el paso del tiempo la población inicial, de tamaño N_0, va enfermando, quedando una población susceptible en un momento t de tamaño N_t. Este número de personas existentes al final del seguimiento de interés depende directamente del número de personas inicial, siendo además función (exponencial decreciente) de la fuerza de morbilidad —representada por la tasa de incidencia— y del tiempo:

Figura 2-4. Gráfico de los susceptibles de enfermar de una cohorte (población fija) en el tiempo



$$N_t = N_0 \exp(-I\Delta t) \; *$$

$$N_t / N_0 = \exp(-I\Delta t) \qquad [1]$$

[*: exp (x) es otra forma de expresar e^x (número e: 2,718... base de los logaritmos naturales)]

Por otra parte la incidencia acumulada o riesgo de enfermar en un periodo Δt (de t_0 a t) es:

$$IA = \frac{N_0 - N_t}{N_0} = 1 - \frac{N_t}{N_0} \quad ;$$

por tanto, sustituyendo con [1]:

$$IA = 1 - \exp(-I\Delta t)$$

Para valores bajos (< 0,1) de $I\,\Delta t$ se puede aproximar a:

$$IA \approx I\Delta t$$

Se ha asumido que la tasa de incidencia es constante en el periodo. Si no se puede asumir esto (normalmente si los periodos son muy largos la poblaciones envejecen, estando por tanto sujetas a riesgos crecientes) conviene realizar estimaciones parciales del riesgo en subperiodos (Δt_i) más cortos. Si se dispone de estimaciones de tasas de incidencia de los subperiodos la fórmula anterior se generaliza a:

$$IA = 1 - \exp\left(-\Sigma\, I_i\, \Delta t_i\right)$$

Ejemplo: En el seguimiento realizado a 1.101.669 mujeres durante 19 años (Pollán y Gustavsson) se obtuvieron las siguientes tasas de incidencia de cáncer de mama por edad:

Edad	Tasa (año^{-1})
25-44	0,00052
45-59	0,00166
60-79	0,00238

El riesgo de desarrollar cáncer de mama en 50 años de una mujer con 30 años será:

$$1 - \exp -[(0{,}00052 \times 15) + (0{,}00166 \times 15) + (0{,}00238 \times 20)] = \mathbf{8\%.}$$

1.3. Prevalencia

La prevalencia se define como la *proporción* de sujetos con una determinada característica, normalmente enfermedad, en un momento o periodo.

$$P = \frac{\text{Número de personas enfermas}}{\text{Total población}} .$$

Se trata de una medida con un carácter estático. Es como una fotografía dónde se refleja la magnitud de un problema en una población en un momento dado. Normalmente se utiliza la *prevalencia puntual*, cuando se hace referencia a un momento concreto. La *prevalencia de periodo* hace referencia a las personas que en algún momento del periodo correspondiente estuvieron enfermas.

Como proporción puede expresarse en términos de probabilidad. En este caso la probabilidad de *estar* enfermo en un momento.

1.4. Relación entre prevalencia e incidencia

La prevalencia puede considerase como un subconjunto (Figura 2-5) compuesto por personas con la enfermedad; al subconjunto se entra enfermando y se sale por muerte o curación. El flujo de entradas está determinado por la tasa de incidencia de enfermedad (I); el flujo de salida está determinado por la tasa de incidencia de 'terminaciones de enfermedad' (I'). Esta es equivalente al inverso de la duración media de la enfermedad $1/\overline{D}$ (se recuerda que el inverso de una tasa de incidencia es el tiempo medio hasta la ocurrencia). En una situación estable, sin migraciones, el número de entradas al subconjunto de prevalencia es igual al numero salidas. Siendo Cp el nº de casos prevalentes y N el tamaño de la población:

$$\underbrace{(N - Cp) \cdot \Delta t \cdot I}_{\text{entradas}} = \underbrace{Cp \cdot \Delta t \cdot 1/\overline{D}}_{\text{salidas}}$$

$Cp/N–Cp$ es la razón entre los casos y los 'no casos', equivalente a $P/1–P$. Esta medida se denomina *odds de prevalencia*. Por tanto:

$$P/1\text{-}P = I \cdot \overline{D}$$

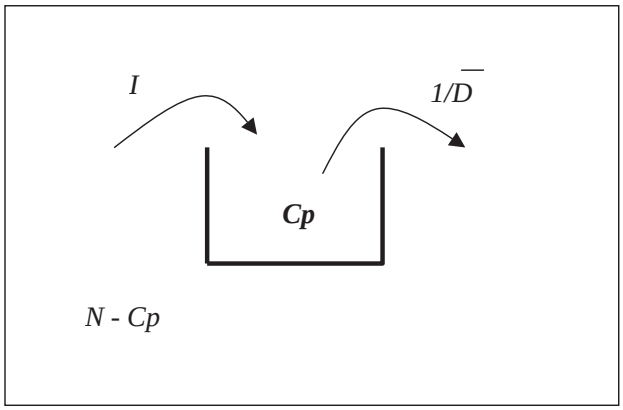
Como P = odds/ 1+odds:

$$P = \frac{I \cdot \overline{D}}{1 + I \cdot \overline{D}} \ .$$

Si la prevalencia es baja (< 0,1): $P \approx I \cdot \overline{D}$

Las medidas de prevalencia suelen tener limitaciones para su utilización en estudios cuyo objetivo es comprobar hipótesis causales (estudios analíticos). Esto se debe a que están, como se ha visto, afectadas por la duración de la enfermedad. Por otra parte se produce el efecto de la modificación de las exposiciones ocasionadas por la propia enfermedad. Por ejemplo, si se quiere estudiar la relación entre ejercicio físico y enfermedad cardiaca, la utilización de casos prevalentes ocasionará que personas con enfermedad presenten valores muy reducidos de ejercicio físico debido a la enfermedad, obteniéndose probablemente una asociación en sentido contrario. Otro ejemplo es la alteración de los niveles de colesterol después de un infarto de miocardio. Si se estudia la relación entre el colesterol y el riesgo de infarto basándose en casos prevalentes se obtendrán estimaciones sesgadas. Se recomienda el libro de Szklo y Nieto para el estudio de estos sesgos.

Figura 2-5. Prevalencia e incidencia. Representación de los flujos de enfermedad de una población estable



1.5. Otras medidas

Odds. Es la razón entre una proporción y su complementaria. Expresa cuánto más probable es la ocurrencia de un fenómeno en relación con la no-ocurrencia. Su rango está entre cero e infinito. Cuanto menor es el valor de la probabilidad más se acerca al valor de la odds. Son muy utilizadas cuando se comparan entre sí mediante razón ('odds ratio').

$$\text{odds} = \frac{p}{1-p} \quad ; \quad p = \frac{\text{odds}}{1+\text{odds}} \ .$$

Tasa de letalidad. Es la proporción de muertes entre los enfermos. Es la incidencia acumulada de muerte en un grupo de enfermos.

Tasa de ataque. Es la proporción de casos que se producen en una población cerrada y en un tiempo muy determinados. Es una incidencia acumulada. Normalmente se utiliza esta terminología en el estudio de brotes epidémicos. La *tasa de ataque secundaria* representa la

probabilidad de enfermar en un sujeto expuesto a un caso primario, siendo por tanto el número de enfermos entre los contactos de un caso.

2. MEDIDAS DE EFECTO

Para poder estimar el efecto que tiene la exposición a un factor de riesgo en la ocurrencia de una enfermedad es necesario conocer la parte de la tasa de incidencia en población expuesta que no se debe a la exposición, es decir, se debe restar *la tasa de incidencia que hubiera tenido esta misma población expuesta caso de no haber estado expuesta*. Medir esto no es posible. Lo que más se puede aproximar a este planteamiento teórico (llamado 'contrafáctico' o 'contrafactual') ocurre en los *estudios cruzados* en los que las mismas personas están expuestas unas veces y no expuestas otras (por ejemplo, a un tratamiento paliativo); aún en estos casos las circunstancias no son las mismas. Se puede obtener una buena aproximación calculando la tasa de incidencia de enfermedad en una población no expuesta muy parecida a la expuesta. En otras palabras, si se pudiera asumir que dos poblaciones son iguales en todo excepto en una variable —exposición o no exposición– las diferencias que se encontraran en sus tasas de incidencia se atribuirían *necesariamente* a la exposición.

2.1. Medidas absolutas. Riesgo atribuible

La *diferencia de tasas* (*DT*) de incidencia representa una medida del *efecto absoluto* atribuible al factor de riesgo. Si las medidas de frecuencia comparadas son incidencias acumuladas se habla de *diferencia de riesgos*. Ambas medidas son denominadas clásicamente como *riesgo atribuible*.

Ejemplo: Deseamos conocer el *efecto* de la exposición a radiaciones en el desarrollo de leucemia. La tasa de incidencia de leucemia en una población expuesta a radiación (I_1) fue de 15,8 por 10^5 personas-año. La tasa equivalente en una población no expuesta (I_0) fue de 3,6 por 10^5 personas-año. El efecto absoluto atribuible a la radiación es:

$$DT = I_1 - I_0 \quad ; DT = 15{,}8\ /10^5/\text{año} - 3{,}6\ /10^5/\text{año} = 12{,}2 \text{ por } 10^5 \text{ personas-año}$$

Si las poblaciones de expuestos y no expuestos no difieren en otros factores relevantes se puede considerar que la radiación es responsable de esa cantidad de casos.

El *rango de valores* que puede tener una diferencia de tasas está entre $-\infty$ y $+\infty$; las unidades de medida son las mismas que tengan las tasas implicadas: año^{-1}, mes^{-1}... Si se trata de una diferencia de riesgos, el rango está entre –1 y +1 y sin unidades ya que es una diferencia de proporciones. El valor que expresa la ausencia de efecto, el llamado *valor nulo*, es 0 para ambas medidas.

2.2. Medidas relativas. Riesgo relativo

El riesgo atribuible es una medida cuya compresión y manejo son importantes en epidemiología y en salud pública pero, como medida absoluta, no expresa adecuadamente la *fuerza de la asociación* si no se relaciona con alguna medida de la magnitud basal de la enfermedad (la incidencia en los no expuestos). Es importante conocer *cuántas veces* más frecuente es la enfermedad en un grupo con respecto al otro. Una medida del *efecto relativo* es:

$$ER = \frac{I_1 - I_0}{I_0} = RT - 1\,.$$

Normalmente se elimina la constante y se utiliza como medida relativa la *razón entre las dos tasas* (*RT*). Siguiendo con el ejemplo anterior:

$$RT = \frac{I_1}{I_0} \quad ; \quad RT = \frac{15{,}8/10^5\ /\text{año}}{3{,}6/10^5\ /\text{año}} = 4{,}4\ .$$

La incidencia en los expuestos supone 4,4 veces la incidencia en los no expuestos al factor de riesgo. También es habitual utilizar el valor del *efecto relativo* (que vale 4,4 – 1 = 3,4) expresado en porcentaje: "La incidencia es un 340% superior en los expuestos con respecto a los no expuestos."

Cuanto mayor es este valor mayor es la fuerza de la asociación entre el factor de riesgo y la enfermedad. Las razones de tasas carecen de unidades y sus posibles valores oscilan entre 0 e infinito.

En este caso el *valor nulo* es 1.

Si el valor es >1 ⟶ la exposición incrementa el riesgo de adquirir la enfermedad.

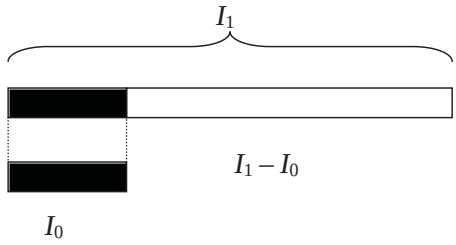
Si el valor es <1 ⟶ la exposición disminuye el riesgo de adquirir la enfermedad (exposición preventiva)

Si en lugar de tasas se utilizan incidencias acumuladas se habla de *razón de riesgos*. Tanto razón de tasas como razón de riesgos son conocidas genéricamente como *riesgo relativo* (*RR*).

2.3. Proporciones atribuibles

La diferencia de tasas (o riesgo atribuible) puede ser expresada como parte del total de la incidencia en expuestos. Tal medida se conoce como *proporción atribuible en expuestos* (PA_E), también *porcentaje de riesgo atribuible* y *fracción etiológica*.

Gráficamente:



$$PA_E = \frac{I_1 - I_0}{I_1}\ .$$

Siguiendo con el ejemplo anterior:

$$PA_E = \frac{15{,}8 - 3{,}6}{15{,}8} = 0{,}77\ .$$

Es decir, el 77% de los enfermos expuestos a radiación deben su enfermedad al hecho de estar expuestos. En otros términos: Si se actúa sobre la población expuesta eliminando tal exposición se previene un 77% de los casos en los expuestos.

Esta proporción expresa la *probabilidad* de que el factor de riesgo sea la causa de la enfermedad en una persona expuesta. Esta probabilidad depende únicamente de la *fuerza de la asociación* (el riesgo relativo):

$$PA_E = \frac{I_1 - I_0}{I_1} = 1 - \frac{1}{RR} = \frac{RR-1}{RR} \ .$$

Conforme aumenta el valor del riesgo relativo la probabilidad se acerca a 1.

Una medida similar puede ser calculada para la población total (expuestos y no expuestos), la *proporción atribuible poblacional* (PA_P):

$$PA_P = \frac{I_T - I_0}{I_T} \qquad [1]$$

Donde I_T es la tasa de incidencia en la población total.

El valor de la PA_P también está en función del riesgo relativo, pero además depende de la cantidad de población que esté expuesta (la *prevalencia de exposición*). Por muy 'patógeno' (*RR* muy alto) que sea un factor de riesgo producirá pocos casos en una población si muy pocas personas están expuestas --y al revés. Dado que la I_T puede expresarse como el promedio ponderado (por la proporciones P_1 y P_0 de cada grupo) de las incidencias en expuestos y no expuestos:

$$I_T = P_1 I_1 + P_0 I_0 \ ,$$

y considerando que $P_0 = 1 - P_1$ podemos sustituir en [1] y simplificando llegar a:

$$PA_P = \frac{RR-1}{RR-1+\frac{1}{P_1}} \ .$$

Si por ejemplo los expuestos suponen el 2% del total de la población:

$$PA_P = \frac{4{,}4-1}{4{,}4-1+\frac{1}{0{,}02}} = 0{,}06 \ .$$

Es decir, en una población conjunta de expuestos y no expuestos el 6% de todos los casos de enfermedad son atribuibles al factor de riesgo.

Este parámetro tiene gran importancia en Salud Pública ya que el número de casos atribuibles (potencialmente prevenibles) puede variar mucho en función del grado de exposición de la población. Por ejemplo, un factor de riesgo cuya asociación con la enfermedad sea débil (*RR* bajo) puede ser responsable de un número de casos mayor que un factor con una asociación fuerte, si la exposición está suficientemente extendida en la población.

Ejemplo: Queremos evaluar el efecto de otro factor de riesgo (FR) para leucemia. La fuerza de la asociación es bastante menor que en el caso de radiación: *riesgo relativo de 1,5*. La *proporción atribuible en expuestos* será:

$$PA_E = \frac{1{,}5-1}{1{,}5} = 0{,}33 \ .$$

Considerablemente más baja que en el caso de la exposición a radiación. Sin embargo, el FR está presente en gran parte de la población: el 40% de la población está expuesta al FR (P_1 = 0,4) por lo que la *proporción atribuible poblacional* es:

$$PA_P = \frac{1{,}5-1}{1{,}5-1+\frac{1}{0{,}4}} = 0{,}16 \ .$$

En contraste con el 6% atribuible a radiación, el segundo FR es responsable del 16% de los casos.

Si la exposición es *preventiva* ($I_1 < I_0$) se puede calcular la *proporción prevenida en expuestos* (PP_E), que es la proporción de casos prevenidos del total de casos prevenibles por la exposición. Se calcula de forma similar a la PA_E:

$$PP_E = \frac{I_0 - I_1}{I_0} = 1 - RT \ .$$

Si la población expuesta tuviera una incidencia igual que la no expuesta, tendría que haberse registrado un número de casos mayor que el actual. El nº de casos que fueron prevenidos por la exposición es la diferencia entre este nº de casos potenciales y los casos actualmente registrados; la parte que suponen los casos prevenidos del total de casos potenciales es la *proporción prevenida* por la exposición.

Ejemplo: Un hipotético estudio de intervención trata de determinar el efecto que tiene una nueva vacuna de varicela. Finalizado el seguimiento las *incidencias acumuladas* son:

$$IA_1 = \frac{9}{1.200} = 0{,}0075 \ ; \ IA_0 = \frac{130}{2.000} = 0{,}065$$

$$RR = \frac{0{,}0075}{0{,}065} = 0{,}115 \ .$$

Es decir, los expuestos a la vacuna tienen un riesgo menor de enfermedad que los no expuestos. Si la población expuesta abandonara tal exposición se registrarían un número adicional de casos hasta contabilizar una incidencia acumulada igual a la de la población no expuesta:

$$\frac{\text{nº de casos prevenibles}}{1200} = 0{,}065 \ ;$$

nº casos prevenibles = 78.

La diferencia entre los *casos ocurridos* (9) y los *potencialmente prevenibles* por la exposición es el nº de casos que han sido prevenidos: 78 – 9 = 69.

La *proporción prevenida* es: 69/78 = 0,88.

Se obtiene el mismo resultado mediante la fórmula descrita anteriormente:

$$PP_E = 1 - RR = 1 - 0{,}115 = 0{,}88.$$

En estos casos la proporción prevenida se denomina *eficacia vacunal* y suele expresarse en porcentaje (88%). Hay que recordar que estas medidas tienen sentido si se asume que las poblaciones de expuestos y no expuestos difieren poco en otros factores relevantes.

2.4. Medida del efecto en estudios de casos y controles

El diseño de casos y controles no permite que se realice una medición directa de las tasas de incidencia y por tanto no se pueden calcular medidas de efecto como la diferencia de tasas o el riesgo relativo. La estrategia en estos estudios consiste en comparar el nivel de exposición en los enfermos (casos) con el equivalente en un grupo de no enfermos (controles). Una medida que relaciona el grado de exposición en ambos grupos es la *odds ratio* (*OR*) o razón de odds. Se ha demostrado que la *OR* es un buen estimador del riesgo relativo.

		Casos	Controles
Exposición	Si	*a*	*b*
	No	*c*	*d*

La *odds de exposición* entre los casos es:

$$\frac{\text{probabilidad de estar expuesto}}{\text{probabilidad de no estar expuesto}} = \frac{a/a+c}{c/a+c} = \frac{a}{c} \ ;$$

y la *odds de exposición* entre los controles:

$$\frac{\text{probabilidad de estar expuesto}}{\text{probabilidad de no estar expuesto}} = \frac{b/b+d}{d/b+d} = \frac{b}{d} \ .$$

Por tanto la *odds ratio*:

$$OR = \frac{a}{c} \Big/ \frac{b}{d} = \frac{a \cdot d}{b \cdot c} \ .$$

El valor de la *OR* se utiliza como medida de magnitud de la asociación exactamente igual que el riesgo relativo ya que se ha demostrado que en la mayoría de las situaciones, es una buena aproximación del mismo. Por tanto, *todas las consideraciones acerca del riesgo relativo son válidas para la OR*: carece de unidades, el rango de valores está entre cero e infinito y el valor nulo que expresa no asociación es 1.

Ejemplo: Los resultados de un estudio que pretende evaluar la asociación entre antecedentes familiares de cáncer colorrectal y cáncer colorrectal son:

		Casos	Controles
Antecedentes familiares	Si	112	108
	No	1.472	2.771

$$OR = \frac{112/1.472}{108/2.771} = \frac{112 \cdot 2.771}{1.472 \cdot 108} = 1{,}95 \ .$$

Este valor es la estimación del riesgo relativo y se interpreta de la misma forma: el riesgo de cáncer colorrectal de las personas con antecedentes familiares es un 95% mayor que el de las personas que carecen de tales antecedentes.

En los estudios de casos y controles al no poder medir incidencias no se puede hallar el riesgo atribuible pero sí se pueden obtener estimaciones de la *proporción atribuible en expuestos* y de la *proporción atribuible poblacional*:

$$PA_E = \frac{OR-1}{OR} \quad ; \quad PA_P = \frac{OR-1}{OR-1+\frac{1}{P_e}} \quad .$$

Siendo *Pe* la proporción de expuestos en la población general (si no se dispone de este dato la proporción de expuestos entre los controles ofrece una buena estimación).

3. REFERENCIAS BIBLIOGRÁFICAS

GORDIS L. *Epidemiology.* Filadelfia: Saunders, 1996.

HENNEKENS CH, BURING JE. *Measures of disease frequency and association.* En: Epidemiology in Medicine. Boston: Little, Brown and Company, 1987.

KAHN HA, SEMPOS CT. *Attributable risk.* En: Statistical Methods in Epidemiology. Nueva York: Oxford University Press, 1989.

KAHN HA, SEMPOS CT. *Relative risk and odds ratio.* En: Statistical Methods in Epidemiology. Nueva York: Oxford University Press, 1989.

KLEINBAUM DG, KUPPER LL, MORGENSTERN H. *Measures of association*. En: Epidemiologic Research. Principles and Quantitative Methods. Nueva York: Van Nostrand Reinhold, 1982.

KLEINBAUM DG, KUPPER LL, MORGENSTERN H. *Measures of disease frequency: Incidence.* En: Epidemiologic Research. Principles and Quantitative Methods. Nueva York: Van Nostrand Reinhold, 1982.

KLEINBAUM DG, KUPPER LL, MORGENSTERN H. *Measures of potential impact and other summary measures.* En: Epidemiologic Research. Principles and Quantitative Methods. Nueva York: Van Nostrand Reinhold, 1982.

MORGENSTERN H, KLEINBAUM D, KUPPER LL. *Measures of disease incidence used in epidemiologic research. Int J Epidemiol* 1980; 9: 97-104.

NIETO GARCÍA FJ, PERUGA URREA A. *Riesgo atribuible: Sus formas, usos e interpretación. Gac Sanit* 1990; 18:112-117.

POLLÁN M, GUSTAVSSON P. *High-risk occupations for breast cancer in the Swedish female working population. Am J Public Health* 1999; 89: 875-881.

ROTHMAN KJ, GREENLAND S. *Measures of disease frequency.* En: Rothman KJ, Greenland S. Modern Epidemiology. Filadelfia: Lippincott-Raven, 1998.

ROTHMAN KJ. *Medidas del efecto.* En: Epidemiología Moderna. Madrid: Díaz de Santos, S.A. 1987.

SZKLO M, NIETO FJ. *Epidemiology; beyond the basics.* Gaithersburg: Aspen, 2000.

SZKLO M, NIETO FJ. *Measuring associations between exposures and outcomes.* En: Epidemiology beyond the basics. Gaithersburg: Aspen, 2000.

III. DISEÑO Y TIPOS DE ESTUDIOS EPIDEMIOLÓGICOS

Jesús Castilla

En un sentido amplio, estudio epidemiológico es cualquier actividad en la que se recurre al método epidemiológico para profundizar en el conocimiento de temas relacionados con la salud. En la práctica, la mayoría de los estudios epidemiológicos tienen como fin aportar información que sirva de apoyo a la toma de decisiones en la planificación o gestión de actividades relacionadas con la salud. No obstante, tienen especial interés aquellos dirigidos a ampliar el conocimiento científico sobre un tema concreto y, cuando cumplen condiciones adecuadas para ello, pueden considerarse verdaderos estudios de investigación.

El método de investigación epidemiológica, como variante del método científico-experimental, consta de las siguientes etapas: observación y descripción de la realidad, elaboración de hipótesis, verificación de la hipótesis, y resolución e inferencia causal.

El paradigma de la investigación es el experimento científico, que consiste en realizar observaciones bajo circunstancias controladas, en las cuales el investigador modifica las condiciones con el fin de establecer el efecto que tiene dicha modificación sobre las observaciones. De forma ideal el experimento compara conjuntos de circunstancias que sólo difieren en la presencia o ausencia del factor objeto de estudio. En la práctica, se considera aceptable si la variabilidad de otros factores distintos del que es objeto de estudio es suficientemente pequeña como para no afectar a los resultados de forma importante.

1. DISEÑO

Cuando hablamos de la arquitectura o diseño de un estudio nos referimos a los procedimientos y métodos por los cuales se seleccionan los sujetos, se recoge y analiza la información y se interpretan los resultados. En otras palabras, el diseño es la estrategia que elegimos para cumplir los objetivos planteados.

Un estudio epidemiológico es un proceso que consta de varias fases: definición del problema de investigación, elección de la estrategia o diseño del estudio y planificación de las actividades.

Toda investigación debe partir del conocimiento profundo y actualizado del tema que va a ser objeto de estudio. La primera fase consiste en definir el problema a investigar, que ha de incluir la justificación del estudio, el propósito general, los objetivos específicos y la definición de la hipótesis operativa, que ha de ser contrastable por métodos epidemiológicos.

En la segunda fase se decide el diseño o estrategia general que se va a emplear para dar respuesta a los objetivos y contrastar la hipótesis de investigación. Esta decisión debe realizarse partiendo de un buen conocimiento de las características y ventajas de los distintos tipos de estudios epidemiológicos, primando aquellos que garanticen una mayor validez y adecuación al objetivo, pero teniendo también en cuenta los recursos humanos, económicos y de tiempo disponibles.

Una vez decidido el objetivo y el diseño se han de planificar las distintas actividades del estudio, que incluyen la definición de la población de estudio, la determinación del tamaño muestral, el método de selección de la muestra, los criterios de inclusión y exclusión, el periodo

de estudio, la selección y definición de las variables, las fuentes de información que se van a utilizar, y cómo se van a desarrollar el procesamiento y análisis de los datos.

2. CRITERIOS DE CLASIFICACIÓN DE LOS ESTUDIOS EPIDEMIOLÓGICOS

El interés de la clasificación de los estudios epidemiológicos es indudable. Proporciona una terminología concisa para la comunicación entre los profesionales y un esquema que facilita la transmisión del conocimiento. A la vez, ayuda a prever los posibles sesgos y limitaciones que pueden aparecer en función del tipo de estudio e identificar el diseño más adecuado para cada objetivo.

Existen muchas clasificaciones de los estudios epidemiológicos, pero la mayoría se basan en algunos criterios básicos comunes: la finalidad del estudio, la unidad de análisis, la direccionalidad, la forma de selección de la muestra, la relación temporal y el control de la asignación del factor a estudio por parte del investigador. A continuación se describen de forma sucinta cada uno de estos criterios.

2.1. Finalidad

Los estudios epidemiológicos se clasifican según su finalidad en descriptivos y analíticos. Los *estudios descriptivos* son los que estudian la frecuencia y distribución de los fenómenos de salud y enfermedad, mientras que los *analíticos* se dirigen a evaluar presuntas relaciones de causa-efecto. En otras palabras, los estudios descriptivos tratan de dar respuesta a preguntas sobre el ¿dónde?, ¿cuándo?, ¿quiénes? y ¿cómo?; mientras que los estudios analíticos tienen por objeto responder al ¿por qué? de los fenómenos de salud y enfermedad (Figura 3-1).

Entre la epidemiología descriptiva y analítica existen múltiples conexiones. Aunque la observación y descripción de la realidad en muchos casos es un fin en si mismo, también puede ser el primer paso del método científico experimental. En los estudios analíticos es frecuente que haya un componente descriptivo inicial.

Entre los estudios descriptivos suelen incluirse también aquellos que exploran presuntas relaciones de causa-efecto, pero que debido a sus limitaciones en el diseño, sus resultados no son aceptados como concluyentes. Algunos autores se refieren a estos estudios como diseños observacionales incompletos o de cribado de hipótesis.

2.2. Unidad de análisis

Todos los estudios epidemiológicos tienen en común analizar problemas de salud en la población; no obstante, la unidad de análisis puede ser:

— *el individuo,* cuando se utilizan datos desagregados de cada sujeto, o

— *poblaciones,* cuando se utilizan datos agregados de grupos de individuos (países, regiones, ciudades, distritos, familias, colegios, empresas, etc). Estos suelen llamarse estudios ecológicos.

Algunos estudios no se quedan en el nivel del individuo y tienen en cuenta la distinta situación que puede tener cada sujeto respecto al tema de estudio en distintos periodos de tiempo.

2.3. Direccionalidad

Kramer y Boivin proponen clasificar los estudios según su direccionalidad atendiendo al orden en que se investiga la asociación entre la causa y el efecto. Consideran tres posibilidades:

— Hacia delante, o *secuencia desde la causa al efecto.* Se seleccionan los sujetos en función de la exposición, y tanto en las personas expuestas como en las no expuestas se evalúa la aparición del efecto. Esta es la característica principal de los estudios de cohortes.

— Hacia atrás, o *secuencia desde el efecto a la causa.* Se seleccionan personas, unas que presentan el efecto y otras no, y se evalúa la presencia de la exposición en todos. Un ejemplo de ello serían los estudios de casos y controles.

— Simultánea o *sin direccionalidad.* La exposición y el efecto son evaluados a la vez. Estos serían los *estudios transversales.*

Las dos primeras situaciones, en las que sí hay direccionalidad, se engloban bajo el término de *estudios longitudinales* y permiten establecer el orden en el que se producen los acontecimientos en el tiempo, que es un argumento fundamental para demostrar relaciones de tipo causal. No obstante, este argumento sólo puede considerarse cumplido cuando el tiempo entre la exposición y el efecto tiene en cuenta el periodo de inducción correspondiente.

2.4. Selección de la muestra

Según el criterio utilizado para elegir a partir de la población diana a los sujetos que conformarán la muestra a estudiar, existen dos posibilidades:

— Muestreo representativo. Consiste en seleccionar una muestra representativa de la población diana.

— Muestreo de conveniencia. Cuando el efecto que se quieren estudiar es poco frecuente puede ser más eficiente muestrear personas que presentan el efecto, y después, buscar el grupo de comparación más adecuado. Por el contrario, cuando la exposición a estudiar es rara, puede interesar ir a buscar directamente personas expuestas y compararlas con un grupo de personas no expuestas. Esta última situación se produce en el estudio de exposiciones que ocurren en determinadas actividades laborales.

2.5. Relación temporal o proximidad

El tiempo transcurrido desde que se produjeron los hechos que se analizan hasta el momento en el que se realiza el estudio puede influir en la validez de la información. En función de esta relación temporal los diseños se clasifican en:

— *Históricos o retrospectivos.* Estudian hechos ocurridos antes del comienzo del estudio. La información puede obtenerse a partir de registros existentes, como por ejemplo en el caso de los estudios de cohortes retrospectivas, o indagando en las entrevistas sobre hechos ocurridos en el pasado.

— *Concurrentes o prospectivos.* Consideran únicamente eventos que se producen a partir del momento de inicio del estudio.

— *Mixtos.* Estudian tanto hechos históricos como concurrentes.

Conviene no confundir la relación temporal y la direccionalidad. Así por ejemplo, un estudio de casos y controles puede ser prospectivo, si se realiza sobre casos incidentes, o retrospectivo, si incluye casos prevalentes.

2.6. Control de la asignación de los factores de estudio

Los *estudios experimentales* son aquellos en los que el investigador controla la asignación de la exposición a estudio. Todos ellos son de tipo analítico, y por razones éticas, suelen limitarse exclusivamente a evaluar intervenciones terapéuticas o preventivas.

Estudios observacionales son todos aquellos en los que el investigador no controla la asignación de la exposición, limitándose a analizar factores cuya presencia o ausencia en los individuos se ha producido por un motivo independiente a la investigación.

3. CLASIFICACIÓN DE LOS ESTUDIOS EPIDEMIOLÓGICOS

Existen muchas clasificaciones de los estudios epidemiológicos, la mayoría de las cuales se basan en los criterios mencionados en el apartado anterior. En las Tablas 3-1, 3-2 y 3-3, y en la Figura 3-1 se muestran varios ejemplos. Las principales diferencias entre las diversas clasificaciones radican en la importancia y prioridad que dan a cada criterio y en discrepancias en la terminología utilizada. Esto pone de manifiesto la dificultad para establecer un esquema que permita clasificar la enorme diversidad de estudios posibles. Por otra parte, no es raro que en una misma investigación haya más de un objetivo y coexistan varios diseños diferentes. Además, existen diseños mixtos, híbridos y situaciones intermedias que rompen la generabilidad de las clasificaciones.

3.1. Estudios descriptivos

Los estudios descriptivos son los que tienen por objeto observar y describir la realidad en función de las variables de tiempo, lugar y persona. Son con diferencia los estudios más frecuentes y la aplicación más habitual de la epidemiología.

La gran mayoría de estos estudios se realizan con la finalidad de aportar información sobre aspectos concretos de la realidad, que sirva de apoyo a la toma de decisiones en los distintos ámbitos relacionados con la salud: actividad asistencial, planificación sanitaria, gestión de recursos de salud, intervenciones en salud pública y políticas de salud. En la medida que los estudios descriptivos suponen la observación y descripción de la realidad, ofrecen también una fuente de hipótesis para la investigación epidemiológica.

Los estudios descriptivos pueden utilizar información existente (fuentes secundarias) o plantearse la recogida de información a tal efecto. Entre las posibles fuentes de datos están las historias clínicas, cuestionarios específicos, resultados de analíticas de laboratorio, sistemas de vigilancia, estadísticas vitales, bases de datos de actividad asistencial, registros de morbilidad, datos de consumos de recursos sanitarios, etc.

Algunos estudios descriptivos tienen como unidad de análisis el individuo, entre los que están la descripción de un caso, la serie de casos y los estudios transversales. Otros tienen como unidad de análisis la población o el grupo de individuos y son los llamados estudios ecológicos.

Tabla 3-1. Clasificación de los estudios epidemiológicos.

- Estudios descriptivos
  - Estudio de una serie de casos
    - Transversal
    - Longitudinal: análisis descriptivo de una cohorte
  - Estudios descriptivos de datos agregados
    - Análisis geográficos
    - Análisis de series temporales
    - Análisis en función de otras variables
  - Estudios descriptivos de prevalencia
- Estudios observacionales de cribado de hipótesis. Diseños incompletos
  - Estudios ecológicos de correlación
  - Diseños proporcionales
    - Estudios de mortalidad proporcional
    - Estudios de morbilidad proporcional
- Estudios analíticos observacionales
  - Estudios transversales
  - Estudios de casos y controles
    - Con casos prevalentes
    - Con casos incidentes
  - Estudios de cohortes
    - Cohortes históricas (retrospectivas)
    - Cohortes concurrentes (prospectivas)
    - Cohortes mixtas
  - Diseños híbridos
    - Estudio de cohorte-casos
    - Estudio de casos y controles anidado en una cohorte
- Estudios experimentales
  - Estudios cuasi-experimentales o de intervención no aleatorizados
  - Ensayo controlado y aleatorizado
    - Ensayo clínico
    - Ensayo de campo
    - Ensayo comunitario

Tabla 3-2. Clasificación de los estudios observacionales (modificada a partir de Kleinbaum y cols.)

Diseños básicos
- Cohortes
- Caso-control
- Transversales

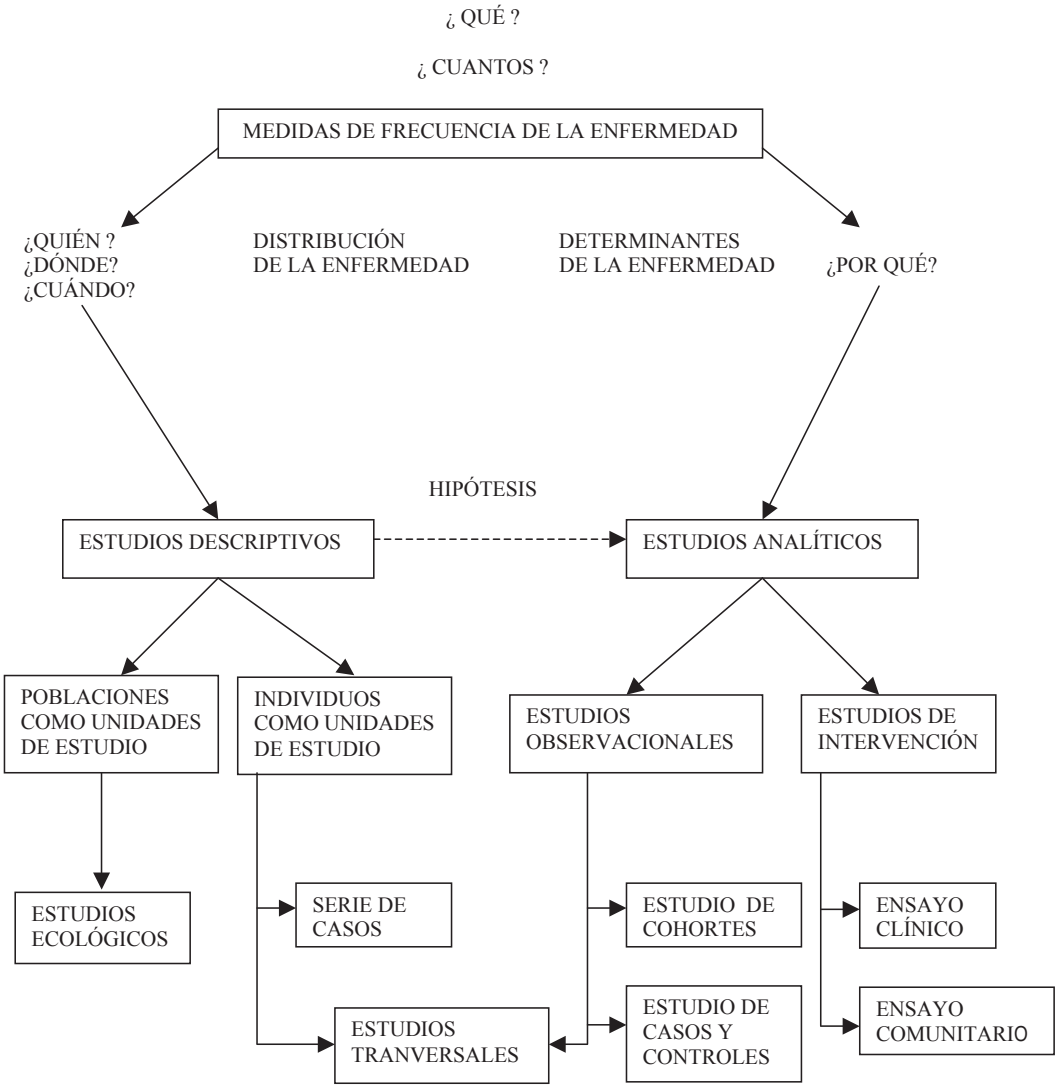
Diseños incompletos
- Estudios ecológicos
- Estudios de mortalidad o morbilidad proporcional
- Estudios de agregación temporo-espacial

Diseños híbridos
- Estudios de casos y controles anidados en una cohorte
- Estudios de caso-cohorte

Tabla 3-3. Clasificación de los estudios epidemiológicos en función del control de la exposición,el objetivo y la unidad de análisis.

Según el control de la exposición	**Objetivo**	**Unidad de análisis**	
		Individuo	**Población**
Experimentales	**Analíticos**	Ensayo clínico Ensayo de campo	Ensayo comunitario
Observacionales	**Analíticos** (comprobación de hipótesis)	Cohortes Casos y controles Transversales analíticos	
	Diseños incompletos (cribado de hipótesis)	Estudios de mortalidad o morbilidad proporcional	Estudios ecológicos de correlación
	Descriptivos	Encuestas de prevalencia descriptivas Series de casos	Estudios descriptivos con datos agregados

Figura 3-1. Clasificación de los estudios epidemiológicos.



Esquema adaptado a partir de materiales docentes de J. M. Martín-Moreno.

3.1.1. Descripción de un caso o de una serie de casos

Describen la experiencia de un único paciente o de un grupo de pacientes con diagnóstico similar, que presentan alguna característica o curso clínico inusual. Pueden ser el punto de partida para una nueva hipótesis. Estos estudios pueden generarse a partir de la actividad asistencial o de la notificación a los sistemas de vigilancia epidemiológica. El paso fundamental en ellos es establecer la definición de caso y los criterios de inclusión y exclusión. Su principal limitación es la falta de un grupo control, lo que impide evaluar hipótesis de tipo causal.

Estos estudios suponen una tercera parte de las publicaciones biomédicas. En muchos casos quedan en la bibliografía como la descripción de una rareza, pero en algunas ocasiones han sido el primer paso para la identificación de una enfermedad emergente o para un nuevo avance científico. La descripción de una serie de casos puede ser la primera constancia de un brote epidémico o de una nueva epidemia. Un ejemplo de ello fue la descripción en 1981 en Los Ángeles de 5 casos de neumonía por *Pneumocystis carinii* en hombres homosexuales previamente sanos. Esta serie de casos fue la primera descripción del sida y el primer paso para el descubrimiento del virus de la inmunodeficiencia humana. Otro ejemplo fue la descripción en España de un aumento de casos de neumonía atípica que no respondían al tratamiento y que más tarde permitió la identificación del síndrome por consumo del aceite tóxico.

La descripción de un caso o serie de casos puede plantearse de forma transversal o longitudinal. Series longitudinales son la cohorte de sujetos a riesgo con el fin de cuantificar la incidencia de una enfermedad, o la cohorte de sujetos enfermos para describir su historia natural.

3.1.2. Estudios descriptivos con datos agregados

También son llamados *estudios ecológicos descriptivos.* Describen la frecuencia de un determinado evento relacionado con la salud en la población, su tendencia en el tiempo, su distribución en el espacio o en función de otras características de las personas como el sexo, edad, profesión, nivel educativo, clase social, etc. Son abordajes muy frecuentes en el campo de la salud pública, como apoyo a actividades de planificación y gestión sanitarias.

Las posibilidades de estos estudios están condicionadas a la existencia de sistemas de información y a la calidad de sus datos. Las fuentes de información pueden ser las estadísticas de mortalidad, los registros de morbilidad, los sistemas de vigilancia epidemiológica u otros sistemas de información sanitaria. Permiten obtener medidas de frecuencia absolutas y relativas, de incidencia, prevalencia o mortalidad, pudiendo establecer comparaciones entre grupos. Los cambios en la frecuencia del evento estudiado a lo largo del tiempo, las diferencias geográficas o las diferencias en función de las características de las personas, son fuentes de hipótesis para estudios analíticos posteriores.

El planteamiento general de estos estudios es sencillo y requiere pocos medios y tiempo para su ejecución. Sus aplicaciones prácticas están experimentando un rápido crecimiento al disponerse de mejores sistemas de información sanitaria y herramientas metodológicas. Los estudios descriptivos en función de la variable tiempo han encontrado un amplio desarrollo a través de los *análisis de series temporales* y de *edad-periodo-cohorte.* Los estudios en función de la variable espacio han expandido sus posibilidades gracias a los *sistemas de información geográfica*.

En este apartado también se incluyen los estudios que exploran posibles agregaciones temporales y/o espaciales de casos. Cuando se detecta una agregación de casos de una enfermedad en el tiempo o en el espacio, por encima de lo esperable por el azar, es muy

sugerente de un foco etiológico común, probablemente infeccioso o ambiental. La detección de una agregación de casos puede aportar hipótesis para la investigación etiológica, y puede ayudar a identificar el foco común para actuar sobre él y controlar la aparición de nuevos casos.

3.1.3. Estudios descriptivos de prevalencia

Las encuestas de prevalencia, también llamadas estudios transversales, pueden realizarse con una finalidad analítica, pero en muchos casos se limitan a objetivos meramente descriptivos. Por ello, dependiendo de cada caso, podrían ser clasificados como estudios descriptivos o analíticos, e incluso algunos autores han llegado a incluirlos en una categoría intermedia de estudios mixtos. Su característica principal es que no tienen direccionalidad, es decir, toda la información recogida hace referencia a un mismo momento en el tiempo. Con frecuencia la selección de la muestra se hace de forma representativa de la población diana, lo que facilita el que los resultados sean generalizables a dicha población. Ofrecen una visión instantánea de lo que ocurre en la población en un momento dado. La medida de frecuencia que proporcionan es la prevalencia.

A este tipo de estudios pertenecen las encuestas que se realizan como apoyo a las actividades de planificación sanitaria de diversa índole, como las encuestas de salud, las encuestas de satisfacción de los usuarios, etc. En ocasiones las variables de estudio son resultados de determinaciones de laboratorio, como es el caso de las *encuestas seroepidemiológicas,* cuyo objeto es conocer la prevalencia de marcadores serológicos de distintas infecciones en la población diana.

Entre los estudios descriptivos transversales se incluyen los que evalúan métodos diagnósticos, comparando los resultados de una técnica con los de la técnica de referencia, y también, los estudios que evalúan la concordancia entre técnicas o entre observadores.

3.2. Diseños observacionales incompletos o estudios de cribado de hipótesis

En este apartado nos referimos a algunos estudios epidemiológicos que se dirigen a explorar presuntas relaciones de causa-efecto, si bien, debido a las limitaciones de su diseño, sus resultados no son aceptados como concluyentes. Entre éstos podrían incluirse los estudios ecológicos y los de morbilidad o mortalidad proporcional. Muchos manuales de epidemiología incluyen este grupo de estudios en el apartado de los descriptivos. Sin embargo, Rothman y Greenland se refieren a éstos como estudios de cribado de hipótesis, ya que cumplirían la función de preseleccionar hipótesis de trabajo para posteriores estudios analíticos. Kleinbaum y colaboradores utilizan el término de diseños observacionales incompletos.

3.2.1. Estudios ecológicos de correlación

Los estudios ecológicos se caracterizan porque las unidades de análisis son grupos de individuos o poblaciones. Utilizan una medida de la frecuencia de la enfermedad en cada una de las poblaciones de estudio y evalúan su relación con algún factor de interés como la edad, la fecha de estudio, la utilización de servicios de salud, el consumo de determinados alimentos, de medicamentos o de otros productos. La medida de asociación que se obtiene es el coeficiente de correlación y por ello también se les llama estudios de correlación.

Los grupos de individuos que se utilizan como unidad de estudio son habitualmente los habitantes de distintos países, regiones o ciudades, pero también pueden compararse cohortes de nacimiento o grupos más pequeños como colegios, empresas o familias. Cuando se compara la población de un mismo lugar en distintos momentos del tiempo se habla de *estudios de tendencia ecológica.*

Un ejemplo de estudio ecológico es el análisis de la correlación a nivel de países entre el consumo per cápita de cigarrillos y la tasa de mortalidad por enfermedad coronaria.

La principal ventaja de estos diseños es que se realizan sobre fuentes de información secundarias, ya disponibles, lo que hace que sean rápidos de ejecución y económicos. Su principal limitación es la imposibilidad de establecer la relación entre la exposición y el efecto a nivel de los individuos. Puede ocurrir que se observe una asociación cuando se analizan poblaciones, sin que exista realmente en el nivel de los individuos, situación que se denomina falacia ecológica. La segunda limitación importante es la dificultad para controlar el efecto de potenciales factores de confusión, porque no siempre se dispone de la información sobre estos factores y además puede resultar imposible separar el efecto de variables que se correlacionan muy estrechamente a nivel poblacional.

3.2.2. Estudios de morbilidad o mortalidad proporcional

Son estudios que tratan de evaluar la asociación entre un factor y un evento, que puede ser la morbilidad o la mortalidad por una determinada causa. Se clasifican en:

— *Estudios de mortalidad proporcional,* si el efecto es la mortalidad.

— *Estudios de morbilidad proporcional,* si el efecto es la incidencia de una enfermedad.

Se realizan a partir de información de registros de mortalidad, o de registros de morbilidad como los de cáncer, de sida, de enfermedades de declaración obligatoria, de altas hospitalarias, etc. En un estudio de mortalidad proporcional la comparación se establece entre la proporción de las muertes en sujetos expuestos que es debida a una determinada causa, con la misma proporción en los sujetos no expuestos. Para algunos autores los estudios proporcionales son un tipo particular de estudios de casos y controles (Rothman y Greenland, 1998), donde los controles son enfermos o fallecidos por otras causas.

La aplicabilidad de estos estudios está limitada en función de las variables que se recogen en los registros. La medida de efecto que se obtiene es la razón de mortalidad (o morbilidad) proporcional. Esta medida es un buen estimador de la razón de tasas cuando se cumple la asunción de que la morbilidad o mortalidad por las restantes causas es independiente de la exposición; no obstante, esta asunción es difícil de demostrar. Otra limitación de los estudios de mortalidad proporcional es que no permiten diferenciar si la exposición se asocia a la enfermedad o a la evolución fatal en los que la padecen.

3.3 Estudios analíticos observacionales

Los estudios analíticos son aquellos que reúnen condiciones adecuadas para evaluar hipótesis y responder al por qué de los fenómenos de salud y enfermedad. En función de la actitud que adopta el investigador los estudios analíticos se clasifican en dos grandes grupos: los estudios observacionales y los experimentales. En los observacionales el investigador se limita a comparar sujetos en función de su exposición al factor objeto de estudio, sin haber intervenido en dicha exposición.

La mayor parte de los autores distinguen tres tipos de *diseños observacionales básicos:* los de corte transversal, los de casos y controles, y los de cohortes. Además hay diseños híbridos que combinan elementos de distintos diseños básicos. En la figura 3-2 se muestran de forma esquemática un ejemplo de los tres diseños básicos y en la tabla 3-4 un análisis comparativo de las ventajas e inconvenientes de los principales diseños analíticos.

3.3.1. Estudios de corte transversal

Los estudios de corte transversal o de prevalencia se utilizan en ocasiones para evaluar hipótesis de tipo causal. La selección de la muestra suele ser representativa de la población de estudio, lo que facilita que los resultados sean generalizables a la población. Se caracterizan por no tener direccionalidad: tanto la evaluación de la exposición como la del efecto hacen referencia a un mismo momento en el tiempo (Figura 3-2). Esto impide establecer la relación temporal entre ambos, limitando la aportación de este diseño en la investigación de relaciones de tipo causal. Otros posibles problemas son que ante una asociación puede no ser fácil distinguir entre un factor de riesgo y un factor pronóstico, y que los casos con mayor duración tienden a estar sobrerrepresentados. Por todos estos motivos se atribuye a los estudios transversales menor valor que a los otros diseños analíticos a la hora de evaluar asociaciones causales. No obstante, los estudios transversales pueden ser adecuados para evaluar exposiciones que no cambian en el tiempo, como el tipo sanguíneo, o cuando la exposición actual es un buen indicador de la exposición relevante respecto al posible efecto, lo cual es más probable si la exposición es constante en el tiempo. En algunos temas, la investigación etiológica suele hacerse sobre casos prevalentes por la dificultad de establecer el comienzo de los procesos, como ocurre con las enfermedades congénitas, la hipertensión arterial, la diabetes, etc.

3.3.2. Estudios de casos y controles

En los estudios de casos y controles se seleccionan dos grupos de sujetos, unos con el efecto que se estudia (casos) y otros sin él (controles). La direccionalidad del estudio se establece hacia atrás, ya que en un segundo paso se indaga en cuáles de los sujetos estuvo presente la exposición que se estudia (Figura 3-2).

En función de la relación temporal puede ser:

— *retrospectivo o histórico,* si se incluyen casos prevalentes,

— *prospectivo o concurrente,* si se incluyen sólo casos incidentes, y

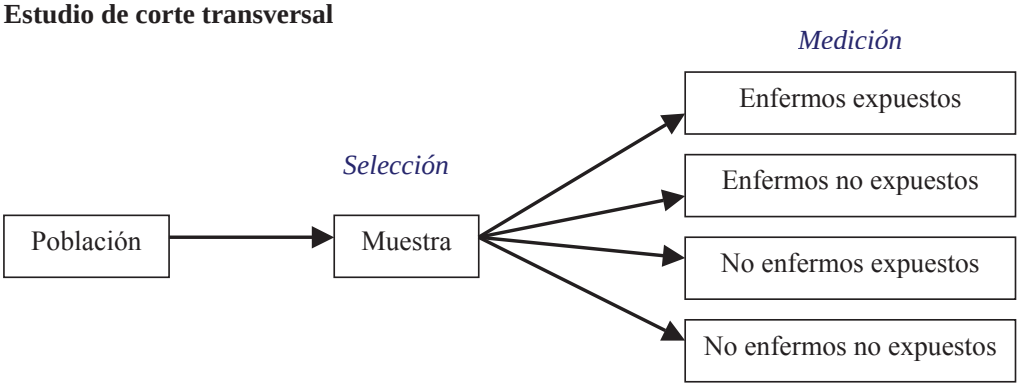
— *mixto*, si se incluyen ambos tipos de casos.

Los casos prevalentes están más expuestos a posibles sesgos de superviviencia selectiva y de memoria. En los estudios de casos y controles no es posible obtener medidas de incidencia y se recurre a la *odds ratio* como medida de asociación que estima la razón de tasas de incidencia.

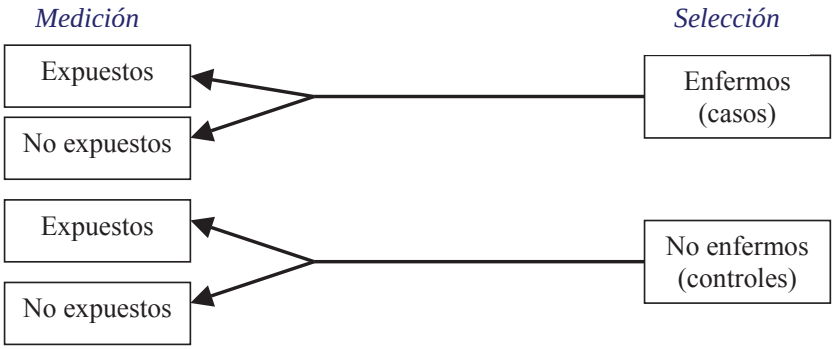
Los estudios de casos y controles son especialmente eficientes para el estudio de enfermedades poco frecuentes o con largos periodos de latencia.

Figura 3-2. Esquema de un ejemplo de diseños de corte transversal, de casos y controles y de cohortes.

Estudio de corte transversal



Estudio de casos y controles



Estudio de cohortes

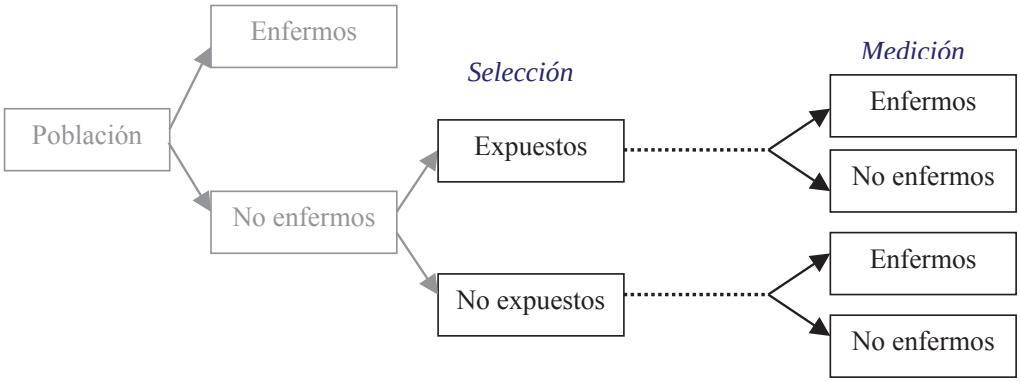


Tabla 3-4. Ventajas e inconvenientes de distintos diseños epidemiológicos analíticos.

	Estudios de corte transversal	Estudios de casos y controles	Estudios de cohortes	Estudios experimentales
VENTAJAS	- Especialmente indicados en enfermedades crónicas - Resultados fácilmente generalizables a la población. - Rápidos de realizar y baratos. - Aportan medida de la prevalencia, útil para tareas de administración sanitaria. - Permiten analizar simultáneamente varios efectos.	- Rápidos de realizar y baratos. - Adecuados para enfermedades con largo periodo de latencia. - Óptimos para enfermedades raras. - Permiten analizar varios factores etiológicos de la enfermedad.	- Óptimos para exposiciones raras. - Permiten estudiar varios efectos de una misma exposición. - Establecen la relación temporal de los acontecimientos. - Permiten medir incidencia. - En las cohortes concurrentes se minimizan los sesgos en la medida de la exposición.	- Permiten establecer la secuencia temporal entre exposición y efecto.. - Permiten un mayor control de la exposición, de sesgos y de potenciales confusores.
DESVENTAJAS	- No permiten establecer la relación temporal entre causa y efecto. - No permiten distinguir entre factores de riesgo y factores pronósticos. - Posibles sesgos en la medida de la exposición. - No útiles en enfermedades raras.	- Ineficientes para exposiciones raras. - No permiten obtener tasas de incidencia. - Puede ser difícil establecer la relación temporal de los eventos. - Tendencia a sesgos de selección y de memoria.	- Ineficientes para enfermedades raras. - Los prospectivos son costosos y de larga duración. - Las cohortes históricas requieren la existencia de registros disponibles. - La validez puede afectarse por pérdidas de seguimiento.	- Limitados a evaluar medidas preventivas o terapéuticas, por razones éticas. - Menor generabilidad de los resultados por realizarse en poblaciones muy seleccionadas. - Sólo estudian una intervención. - Elevado coste.

3.3.3. Estudios de cohortes

La característica que los distingue es que la direccionalidad va desde la exposición hasta el efecto. Los acontecimientos se estudian en la misma secuencia en la que se produjeron, lo que hace que este diseño sea más intuitivo y fácil de entender. La selección de la muestra puede hacerse de forma representativa de la población diana o en función de la exposición, como por ejemplo en las cohortes ocupacionales. Esta última posibilidad es especialmente útil cuando se estudian exposiciones raras. En función de la relación temporal entre los hechos que se estudian y el momento de realización de la investigación, los estudios de cohortes pueden ser históricos, concurrentes o mixtos.

— Los *estudios de cohortes históricas,* también son llamadas *retrospectivas,* se basan en información recogida en registros.
— Los *estudios de cohortes concurrentes o prospectivos* sólo tienen en cuenta los eventos que se producen a partir del comienzo del estudio.
— Los *estudios de cohortes mixtas* son los que combinan información prospectiva y retrospectiva.

Inicialmente todos los sujetos están libres del efecto y se diferencian según la exposición. A partir de ese momento se requiere contemplar el seguimiento de todos los sujetos durante un intervalo de tiempo suficiente para dar margen al periodo de inducción del efecto que se analiza. Al final del seguimiento se evalúa la aparición del efecto en los dos grupos (Figura 3-2).

Los estudios de cohortes permiten establecer la relación temporal entre la exposición y el efecto, y obtener medidas de incidencia. Además si las cohortes son concurrentes se minimizan los sesgos en la medida de la exposición.

3.3.4. Diseños híbridos

Bajo este epígrafe se incluyen los *estudios de cohorte-casos* y los *estudios de casos y controles anidados en una cohorte.* En ambos se parte de un estudio de cohortes. En el análisis se incluyen todos los casos y, como grupo de comparación, se toma una submuestra de la cohorte. De esta forma se reduce considerablemente el número de sujetos que es necesario estudiar, sin que ello suponga una disminución importante de la potencia del estudio. Estos diseños consiguen mejorar mucho la eficiencia cuando la medición de la exposición o de otras variables de análisis es costosa.

3.4. Estudios experimentales

Los estudios experimentales son aquellos en los que la exposición es asignada por el investigador en función de lo establecido previamente en el protocolo. Cuando la asignación se realiza en función de otros criterios, como por ejemplo, la aplicación de un tratamiento siguiendo criterios terapéuticos, podremos estar ante un estudio observacional, pero no experimental.

Con el fin de garantizar la mayor comparabilidad entre los grupos y evitar sesgos, lo ideal es que la asignación de la exposición se realice de forma aleatorizada, aunque esto no siempre es posible. Muchos autores consideran estudios experimentales exclusivamente a aquellos realizados de forma aleatorizada: ensayos clínicos, ensayos de campo y ensayos comunitarios.

Por razones éticas estos estudios únicamente tienen aplicación para evaluar el efecto de exposiciones potencialmente terapéuticas o preventivas. Además, las alternativas que se comparan deben ser a priori igualmente aceptables en función del conocimiento que se tiene sobre el tema; por ejemplo, el mejor tratamiento existente debe ser la referencia de comparación para cualquier nuevo tratamiento.

3.4.1. Estudios cuasi-experimentales o de intervención no aleatorizados

Se utiliza el término de estudios cuasi-experimentales para referirse a aquellos en los que hay asignación controlada de la exposición sin seguir un procedimiento aleatorio. Las unidades de asignación del factor de estudio pueden ser los individuos o también las poblaciones. Dentro de los estudios cuasi-experimentales también se incluyen los estudios en los que la comparación se establece entre los mismos sujetos antes y después de aplicar la intervención.

Este tipo de estudios se utilizan frecuentemente para la evaluación de la eficacia de actividades preventivas y de educación para la salud.

3.4.2. Ensayo clínico

En los ensayos clínicos, también llamados ensayos aleatorizados controlados, las unidades de estudio son pacientes. Su objetivo suele concretarse en evaluar nuevos tratamientos o estrategias de prevención secundaria. Sobre la población diana se aplican los criterios de inclusión y exclusión, previamente protocolizados, y se obtienen los sujetos elegibles. Existe un acuerdo universal de que la asignación aleatoria del tratamiento es el mejor método para garantizar la comparabilidad inicial de los grupos respecto a todas las características. Cuando es posible, conviene emplear enmascaramiento, es decir, que el paciente, la persona que asigna la exposición y la que evalúa el resultado, desconozcan qué pacientes reciben el tratamiento nuevo y cuáles el tratamiento habitual (o el placebo). Finalizado el seguimiento de los pacientes se procede a comparar los resultados de la intervención entre los grupos.

3.4.3. Ensayo de campo

Los ensayos de campo evalúan medidas de prevención primaria y, por tanto, se aplican en sujetos sanos. Como la incidencia de la mayoría de enfermedades en población sana es muy baja, se requiere que estos estudios tengan un elevado número de participantes, para poder obtener resultados concluyentes. Esto limita la aplicabilidad de los ensayos de campo a medidas de prevención en enfermedades frecuentes o muy graves y hace que sean estudios económicamente muy costosos. En ejemplo sería la evaluación del suplemento vitamínico en población sana para prevenir ciertas enfermedades.

3.4.4. Ensayo de intervención comunitaria

Los ensayos de intervención comunitaria evalúan también medidas de prevención primaria, pero se caracterizan porque la unidad de asignación del factor de estudio y de análisis son grupos de individuos o comunidades. Algunos ejemplos de posibles intervenciones comunitarias son la fluoración de las aguas de abastecimiento público, los programas de educación utilizando medios de comunicación de masas, o la intervención medioambiental en el lugar de trabajo. La asignación puede hacerse de forma aleatoria entre los grupos o comunidades, pero lo más frecuente es que no sea así y en este caso pasarían a ser estudios cuasi-experimentales.

4. ESTUDIOS EPIDEMIOLÓGICOS Y GRADO DE EVIDENCIA CIENTÍFICA

Rara vez se atribuye a un único estudio la evidencia científica definitiva respecto a un tema; sin embargo, son muchos los estudios epidemiológicos que aportan algún grado de evidencia, a favor o en contra de presuntas asociaciones de tipo causal. Aún asumiendo la validez de los estudios, no todos aportan un mismo grado de evidencia. Según Maclure las características de los estudios epidemiológicos que permiten alcanzar una mayor validez son por orden de importancia: la aleatorización y enmascaramiento, la desagregación del análisis a

nivel del individuo, el diseño longitudinal y el planteamiento prospectivo. Por tanto, estas características en los estudios les conferirían un mayor grado de evidencia científica.

En la tabla 3-5 se ordenan los principales diseños en función del grado de evidencia que proporcionan. A los ensayos controlados con enmascaramiento se les atribuye el mayor grado de evidencia científica dentro de la investigación epidemiológica. El resto de ensayos controlados y los estudios de cohortes, siempre que la metodología se haya aplicado correctamente y se descarten sesgos importantes, se les reconoce el muy alto grado de evidencia. A los estudios de casos y controles y transversales, cuando se han descartado sesgos importantes, se les atribuye un grado de evidencia elevado. Por último, a los estudios proporcionales, ecológicos y descriptivos en general, se les atribuye un valor orientativo, pero no concluyente.

Tabla 3-5. Clasificación de los diseños epidemiológicos según el grado de evidencia científica que se les atribuye en la investigación. Ordenados de mayor a menor evidencia.

1. Ensayo controlado aleatorizado con enmascaramiento.
2. Ensayo controlado aleatorizado sin enmascaramiento.
3. Ensayo controlado no aleatorizado, estudio de intervención
4. Estudio de cohortes concurrentes (prospectivas)
5. Estudio de cohortes históricas (retrospectivas)
6. Estudio de casos y controles con casos incidentes
7. Estudio de casos y controles con casos prevalentes
8. Estudio de corte transversal
9. Estudio de mortalidad o morbilidad proporcional
10. Estudio ecológico de correlación

5. REFERENCIAS BIBLIOGRÁFICAS

ARGIMÓN PALLÁS JM, JIMÉNEZ VILLA J. *Métodos de investigación clínica y epidemiológica* (Segunda edición). Barcelona: Harcourt, 1997.

BOLUMAR MONTRULL F, REBAGLIATO RUSO M, TORRES CANTERO AM. *Estrategias de diseño en epidemiología. Tipos de estudios.* En: Piedrola Gil (Ed.) Medicina Preventiva y Salud Pública (10ª Edición). Barcelona: Masson, 2000. pp. 79-86.

FEINSTEIN AR. *Clinical epidemiology: the architecture of clinical research.* Philadelphia: WB Saunders, 1985. pp. 12-24.

GUALLAR EM. *Alternativas al diseño tradicional de los estudios de cohortes: estudios de casos y controles híbridos, y estudios de cohortes y casos.* Rev Salud Pública 1991;2:151-165.

HENNEKENS CH, BURING JE. *Epidemiology in medicine.* Boston: Little Brown and Company, 1987.

KELSEY JL, PETITTI DB, KING AC. *Key methodologic concepts and issues.* En: Brownson RC, Petitti DB (Eds.). Applied epidemiology: theory to practice. New York: Oxford University Press, 1998. pp. 35-48.

KLEINBAUM DG, KUPPER LL, MORGENSTERN H. *Epidemiologic research. Principles and quantitative methods.* New York: Van Nostrand Reinhold, 1982.

KRAMER MS, BOIVIN JF. *Towards an "unconfounded" classification of epidemiologic research design.* J Chronic Dis 1987;40:683-688.

MACLURE M. *Taxonomic axes of epidemiological study designs: a refutationist perspective.* J Clin Epidemiol 1991;44:1045-1053.

ROTHMAN KJ, GREENLAND S. *Types of epidemiologic studies.* En: Rothman KJ, Greenland S. (Eds.) Modern epidemiology (Second edition). Philadelphia: Lippincott Raven Publisher, 1998. pp.67-78.

IV. SESGOS Y FACTORES DE CONFUSIÓN

Fernando Villar Álvarez

Si se entiende que los estudios epidemiológicos son ejercicios de medición cuyo objetivo es estimar valores de frecuencia de enfermedad o medidas de efecto de una exposición sobre una enfermedad, se puede apreciar fácilmente que la meta fundamental será conseguir la exactitud (*accuracy* en inglés) en esa medición: estimar el valor del parámetro estudiado con el menor error posible. Existen dos tipos de error que pueden afectar a la medición: error aleatorio y error sistemático.

1. PRECISIÓN Y VALIDEZ

El error aleatorio (o al azar) produce una desviación del resultado obtenido respecto del verdadero valor en un determinado estudio, pero si el estudio se repitiera un número infinito de veces con la misma metodología y no hubiera sesgos, se obtendría un valor medio igual al valor real. El error aleatorio surge debido a que los estudios se basan en un número limitado de observaciones. Por ello, el error aleatorio disminuye al aumentar el tamaño del estudio. El error aleatorio no es constante en cada medición, ni ocurre siempre en el mismo sentido. De forma aleatoria, se obtienen valores a veces mayores y a veces menores que el valor real. Rothman resume el error aleatorio como esa parte de nuestra experiencia que no podemos predecir.

El error sistemático se produce consistentemente en cada medición y ocurre siempre en el mismo sentido (se obtienen valores siempre mayores o siempre menores que el valor real). Por tanto, el error sistemático también produce una desviación del resultado obtenido respecto del verdadero valor en un determinado estudio. Sin embargo, cuando hay un error sistemático si el estudio se repitiera un número infinito de veces con la misma metodología, se obtendría un valor medio distinto del valor real. Por ello, el error sistemático, a diferencia del error aleatorio, no depende del tamaño del estudio, con lo cual no se reduce aumentando su tamaño. En presencia de error sistemático, el aumento del tamaño del estudio únicamente conseguiría aumentar la precisión de la estimación pero no su validez. Es decir, para un determinado nivel de seguridad, el aumento del tamaño del estudio conseguiría disminuir el intervalo de confianza (aumento de la precisión), pero el valor estimado estaría igual de alejado del valor real (igual error sistemático).

La ausencia de error aleatorio o debido al azar se denomina precisión o fiabilidad. Por su parte, la ausencia de error sistemático se denomina validez.

La precisión se puede valorar examinando los intervalos de confianza de las estimaciones. A mayor precisión, menor amplitud del intervalo de confianza.

Como se ha visto, los estudios epidemiológicos se basan en la medición de variables y estas variables son aleatorias, es decir, su valor está determinado por un fenómeno aleatorio (regido por la casualidad, no determinado completamente por otros factores). Por ello, la principal fuente de error aleatorio es el proceso de selección de los sujetos estudiados, denominado *error muestral*. Se debe entender que en todo estudio epidemiológico hay un error muestral, ya que, tal como señala Rothman, aún en los estudios en que se pueda considerar que no hay en sentido estricto una selección de una muestra o muestreo, como puedan ser los de seguimiento, siempre se deben considerar a los sujetos

estudiados como una muestra de la experiencia biológica de una población más amplia. Habitualmente, las diferencias observadas en los resultados obtenidos en distintos estudios realizados en una misma población se deben al error aleatorio.

Ejemplo: En un estudio de prevalencia de la hipertensión arterial en una zona determinada se extrae una muestra A y se obtiene una prevalencia del 20%. Si se extrajera en esa misma zona una muestra B se podría obtener una prevalencia del 22%. Suponiendo que no hay error sistemático, esta diferencia se debería al error aleatorio. Esta variabilidad disminuirá si el tamaño del estudio aumenta.

Otra fuente de error aleatorio es la propia variabilidad (varianza) del parámetro que se mide.

Teniendo en cuenta que la precisión es la capacidad de obtener valores replicados que difieren poco entre sí. El error aleatorio va a originar una escasa reproducibilidad de los resultados y una imprecisión de los estimadores de la frecuencia de la enfermedad o del efecto de la exposición sobre la enfermedad. Sin embargo, el error aleatorio no afecta en sí a la validez, ya que influye por igual en todos los grupos y subgrupos estudiados, pero sí dificulta la posibilidad de encontrar una asociación entre dos variables por la imprecisión de las medidas.

El error aleatorio es un concepto estadístico, en cuya valoración, lógicamente, también deben considerarse elementos de juicio informado. Sin embargo, en la valoración del error sistemático intervienen fundamentalmente elementos de juicio.

En la Tabla 4.1. se muestran las principales diferencias entre el error aleatorio y el error sistemático.

La validez o ausencia de error sistemático se divide en dos componentes: validez interna (validez de inferencia para las propias personas del estudio) y validez externa o generalización (validez de inferencia a personas que están fuera de la población de estudio).

1.1. Validez interna

La validez interna tiene relación con la capacidad de inferir los resultados desde el grupo de personas estudiadas hasta la población base, población elegible o población de referencia (conjunto de personas de donde proceden directamente las personas incluidas en el estudio). Respondería a la pregunta: ¿los resultados que se observan en el estudio son aplicables a la población base?

Ejemplo: En el estudio de seguimiento de Doll y Hill (1964) sobre la relación del consumo de tabaco y el cáncer de pulmón realizado en los médicos británicos, los resultados obtenidos tendrán validez interna si la relación encontrada en los médicos británicos estudiados es cierta en el conjunto de todos los médicos británicos (población base o elegible).

1.2. Validez externa

La validez externa tiene relación con la capacidad de generalización de los resultados desde el grupo de personas estudiadas hasta la población diana o población general (conjunto de personas que están fuera de la población de estudio y para los que se pretende obtener conclusiones válidas de los resultados obtenidos). Respondería a la pregunta: ¿los resultados qué se observan en el estudio son aplicables a cualquier grupo?

Tabla 4.1.- Principales diferencias entre el error aleatorio y el error sistemático.

	ERROR	
	ALEATORIO	**SISTEMÁTICO (SESGO)**
Causa	- Muestreo - Variabilidad del parámetro	- Diseño, ejecución y análisis: · Selección de los sujetos del estudio · Obtención de la información · Presencia de variables externas distorsionadoras
Disminuye al aumentar el tamaño del estudio	Sí	No
Afecta a	Precisión	Validez
Valoración	Concepto estadístico unido a elementos de juicio informado	Intervienen fundamentalmente elementos de juicio informado

Para poder valorar la validez externa es necesario recurrir a criterios e información externos al propio estudio. Para ello se utilizarán los criterios de causalidad (consistencia, plausibilidad biológica, etc.). En última instancia, la validez de una generalización es una cuestión de juicio informado.

Ejemplo: En el estudio de seguimiento de Doll y Hill (1964) sobre la relación del consumo de tabaco y el cáncer de pulmón realizado en los médicos británicos, los resultados obtenidos tendrán validez externa o podrán ser generalizados si la relación encontrada en los médicos británicos estudiados es cierta en el conjunto de toda la población inglesa o de otros lugares (población diana o general). Es decir, si el consumo de tabaco es un factor de riesgo del cáncer de pulmón en cualquier grupo de personas que fumen.

Es muy importante establecer claramente que en ausencia de validez interna la validez externa de un estudio carece de sentido. La validez interna es un prerrequisito para que pueda darse la validez externa. Por tanto, lo prioritario en un estudio epidemiológico es que tenga validez interna, lógicamente, además también será necesario tener en cuenta la validez externa.

Ejemplo: En el estudio de seguimiento de Doll y Hill (1964) sobre la relación del consumo de tabaco y el cáncer de pulmón realizado en los médicos británicos, si la relación encontrada en los médicos británicos estudiados ni siquiera fuera válida para el conjunto de todos los médicos británicos de donde proceden, difícilmente va a ser válida para otras poblaciones más amplias (población diana o general).

Los errores sistemáticos afectan a la validez interna de un estudio y, por lo tanto, indirectamente también a la validez externa del mismo.

2. SESGOS

2.1. Concepto de sesgo

Una primera aproximación al concepto la puede proporcionar el diccionario de la Real Academia Española, que define el sustantivo sesgo como oblicuidad o torcimiento de una cosa hacia un lado, o en el corte, o en la situación, o en el movimiento. Para precisar el concepto epidemiológico se puede recurrir al Diccionario de Epidemiología, que define el sesgo como "desviación de la verdad de los resultados o inferencias, o los procesos que llevan a esa desviación. Cualquier tendencia en la recogida, análisis, interpretación, publicación o revisión de los datos que pueden conducir a conclusiones que son sistemáticamente diferentes de la verdad".

Se entiende por sesgo cualquier error sistemático que afecte a la validez de un estudio, es decir, que aleje el valor observado del valor real o verdadero. Recuérdese que se denomina validez a la ausencia de sesgo sistemático. Como habitualmente no se conoce el valor real y únicamente se dispone del valor observado, en la mayoría de las ocasiones no se tendrá la certeza de la presencia del sesgo ni de su magnitud. Por ello, sólo será posible prevenir o minimizar los sesgos en el diseño y ejecución del estudio mediante una adecuada selección de los sujetos estudiados, una correcta obtención de la información sobre los mismos, y la prevención en el diseño o control en el análisis de las variables externas distorsionadoras. Además, se podrá argumentar razonadamente acerca de su presencia y efecto sobre el parámetro estimado. Por tanto, en la valoración de un sesgo intervienen fundamentalmente elementos de juicio informado.

Como se ha señalado, el error sistemático, a diferencia del error aleatorio, no depende del tamaño del estudio, con lo cual no se reduce aumentando su tamaño. En presencia de error sistemático, el aumento del tamaño del estudio únicamente conseguiría aumentar la precisión

de la estimación pero no su validez, ya que el valor estimado seguirá estando igual de alejado del valor verdadero.

2.2. Tipos de sesgos

Existen distintos tipos de sesgos que pueden disminuir la validez interna de un estudio, es decir, la inferencia de los resultados obtenidos a la población base. Sackett (1979) ha descrito docenas de posibles sesgos capaces de distorsionar la estimación de una medida epidemiológica, pero para simplificar su exposición se pueden agrupar en tres tipos generales:

— Sesgo de selección.
— Sesgo de información.
— Sesgo de confusión.

El sesgo de selección se puede producir en las fases de diseño y ejecución del estudio debido a una inadecuada selección de los sujetos estudiados. Por su parte, el sesgo de información se puede originar en las fases de diseño y ejecución del estudio debido a una incorrecta obtención de la información sobre las personas estudiadas. Finalmente, el sesgo de confusión aparecerá cuando no se haya evitado en el diseño la presencia de variables externas distorsionadoras o no se haya controlado su efecto en el análisis.

Aunque así definidos, aparentemente están claramente diferenciados, en la realidad, en ocasiones es difícil poder distinguir entre estos tres tipos de sesgos, ya que no siempre aparecen tan claramente delimitados. Hay sesgos de información / selección combinados, que tienen componentes de información y de selección. También, puede haber factores que parecen ser responsables de un sesgo de selección y que, sin embargo, bajo determinadas circunstancias, podrían considerarse como factores de confusión. Se puede dar una guía útil para poder distinguir en la práctica entre el sesgo de confusión y el resto de los sesgos. Un sesgo es de confusión si puede ser controlado en la fase de análisis de los datos. Igualmente, cabe señalar las similitudes entre algunos sesgos de selección y los sesgos de clasificación errónea diferencial, como se verá más adelante.

Por tanto, esta clasificación de los sesgos es una simplificación de la realidad, que se hace con própositos fundamentalmente didácticos. Por este motivo, como señalan Szklo y Nieto, no debe entenderse como una taxonomia rígida y mutuamente excluyente.

Además, de los sesgos producidos en las fases de diseño y ejecución de un estudio, también pueden darse sesgos en la fase de difusión o publicación de los resultados. Es lo que se denomina *sesgo de publicación*. En su sentido más amplio, el sesgo de publicación se puede entender como cualquier influencia que reduzca la cantidad de ciencia válida que aparece en la literatura científica, es decir, cuando lo que se publica no es representativo de lo que se investiga. Habitualmente se debe a la tendencia a publicar más los estudios que obtienen resultados positivos (encuentran diferencias estadísticamente significativas) frente a los negativos (no encuentran diferencias estadísticamente significativas). Esto produce una visión incompleta y sesgada de la evidencia disponible.

El sesgo de publicación no es específico del ámbito sanitario. La mayor parte de la investigación sobre este sesgo se ha hecho en el campo de la psicología y la educación. Fue Sterling quien en 1959 publicó el primer trabajo sobre el posible efecto de las decisiones de las publicaciones. El sesgo de publicación es la primera fuente de sesgo de selección en los metanálisis (véase capítulo 12).

Es de destacar algunas iniciativas para minimizar el sesgo de publicación. En la actualización de noviembre de 2003 de los Requisitos de uniformidad para la redacción y edición de

manuscritos presentados a revistas biomédicas, elaborados por el Comité Internacional de Directores de Revistas Médicas, entre los aspectos editoriales relativos a la publicación en revistas biomédicas, se señala la obligación de publicar estudios con resultados negativos. Se dice que las revistas deberían considerar seriamente para su publicación cualquier estudio que haya sido rigurosamente realizado sobre cuestiones que sean importantes y relevantes para los lectores, tanto si los resultados son negativos (aquellos que permiten de forma convincente aceptar la hipótesis nula) como si son positivos (aquellos que permiten rechazar la hipótesis nula). La publicación de estudios con resultados negativos, en particular, contribuyen a evitar sesgos de publicación. Algunos estudios que pretenden ser negativos son, de hecho, poco concluyentes. La publicación de este tipo de estudios es problemática, ya que añaden poco conocimiento biomédico y consumen recursos de las revistas.

2.3. Efectos de los sesgos

Cualquiera de los sesgos (selección, información y confusión), en función del efecto que producen al distorsionar la estimación de una medida epidemiológica, pueden clasificarse en positivos o negativos. Se denominan *positivos* cuando el parámetro epidemiológico presenta un valor superior al que realmente tiene (sobrestima o exagera el efecto), es decir, lo aleja de la hipótesis nula (RR=1). Se denominan *negativos* cuando el parámetro presenta un valor inferior al real (subestima o atenúa el efecto), es decir, lo aproxima a la hipótesis nula. En función de la magnitud de la distorsión producida por la acción del sesgo, un efecto que existe en la realidad puede llegar a anularse por el sesgo o incluso, si la distorsión es aún mayor, puede llegar a invertirse, cambiando el sentido de la asociación, produciendo una modificación *cualitativa* del efecto. De esta forma, un sesgo sería capaz de hacer que un factor de riesgo apareciera como factor de protección o al revés. Este fenómeno tan peligroso, que en el sesgo de confusión recibe el nombre de "Paradoja de Simpson", afortunadamente es muy poco frecuente y, para que se produzca se requiere que la magnitud del efecto real sea pequeña y que la distorsión producida por el sesgo sea grande. Se suelen preferir los sesgos negativos a los positivos, ya que si a pesar de haber un sesgo negativo se encuentra un efecto debido a una exposición, es que esa asociación realmente existe, aunque sea de una magnitud algo mayor que la observada.

Habitualmente, no se va a tener la certeza de si un sesgo está presente o no, y sólo se podrá argumentar sobre su presencia, sentido y magnitud, con la sola excepción del sesgo de confusión, que a diferencia del resto de los sesgos (selección e información), sí se puede cuantificar y controlar en el análisis.

Este hecho nos pone sobre la pista de una de las ideas clave en el tema de los sesgos. Como se verá más adelante, la presencia de algún tipo de sesgo es prácticamente inevitable en cualquier estudio epidemiológico, aunque es cierto que unos tipos de diseño de estudio son más vulnerables que otros a la presencia de sesgos. Los estudios experimentales, como el ensayo clínico aleatorizado, por el propio hecho de la aleatorización (no confundir con el muestreo aleatorio), que teóricamente controla todos los factores de confusión, conocidos y desconocidos, sobre todo en estudios grandes, son poco susceptibles a que se introduzcan en ellos sesgos de confusión. Dentro de los estudios observacionales, es bien conocido que los estudios de casos y controles son más proclives a la presencia de sesgos que los de cohortes. Hay sesgos que son inherentes al propio tipo de diseño del estudio y no se puede impedir su presencia. Se puede señalar como ejemplo, los estudios de casos y controles que utilicen una pregunta anamnésica para obtener la información sobre la exposición, que casi inevitablemente van a producir un sesgo de información denominado sesgo de memoria. Por ello, la importancia de los sesgos no viene dada tanto por el hecho de que estén presentes o no, ya que casi siempre van a actuar, sino por su magnitud y especialmente cuando el efecto estudiado tiene una magnitud pequeña, porque es entonces

cuando el sesgo puede llevar a extraer conclusiones erróneas. Cuando la asociación estudiada es fuerte, es más difícil que el sesgo sea capaz de justificar completamente esa asociación. Téngase en cuenta a este respecto que uno de los criterios de Austin Bradford Hill (1965) para valorar la posible causalidad de una asociación es la fuerza o magnitud del efecto. A pesar de ello, una asociación débil puede ser causal, ya que los criterios de Hill, y el de fuerza también, pueden tener excepciones. La valoración de la magnitud de la asociación es una cuestión de juicio informado, al que ayuda el conocer lo consistentemente grande que es el efecto en distintos estudios sobre la misma relación. La consistencia es otro de los criterios de causalidad y, para algunos autores, es el más importante, ya que es poco probable que distintos estudios sobre una misma relación, que tendrán sesgos diferentes, obtengan un mismo efecto si este no existe realmente. El tema de causalidad será expuesto detalladamente en el Capítulo 13, que trata de la Inferencia Causal en Epidemiología.

2.4. Diferencias entre los sesgos de selección e información y el sesgo de confusión

Aunque, tal como se ha señalado, el sesgo de confusión es un sesgo, tiene algunas particularidades que lo diferencian del resto de los sesgos de selección e información. Por ello, es frecuente que cuando se mencione simplemente sesgo, se esté refiriendo exclusivamente a los sesgos de selección e información. Y que cuando se refiera al sesgo de confusión se mencione simplemente confusión. Para tener siempre presente que la confusión es un sesgo, es decir, es un error sistemático, es bueno anteponer al término de confusión la palabra sesgo.

Es necesario mencionar en este punto que hay autores que no comparten este criterio. Por ejemplo, en los libros de Lilienfeld y Stoley, de Gordis, y de Szklo y Nieto se diferencia entre asociaciones “espurias” debidas a sesgo y asociaciones “indirectas” (o “estadísticas”) debidas a la confusión. Por tanto, conceptualizan la confusión como algo diferente a un sesgo, por cuanto aunque en la confusión la relación no sea causal, sí es real. Sin embargo, aquí se seguirá el criterio propugnado, entre otros, por Rothman y Greenland, que consideran que la confusión es un sesgo, ya que una asociación confundida puede entenderse como un estimado sesgado de una asociación causal.

Los sesgos de selección e información son un problema de diseño y ejecución del estudio que da lugar a una estimación errónea, distorsionada del efecto. Pueden aumentar, disminuir o anular el efecto real. Por tanto, afectan a la validez del estudio. Habitualmente, no se pueden cuantificar ni se pueden controlar en la fase de análisis. Por tanto, deben evitarse mediante un diseño cuidadoso. Frecuentemente son el punto débil del estudio.

Por su parte, la confusión también es un sesgo, pero en este caso está producido por la presencia de una variable externa, denominada factor de confusión, que da lugar a una mezcla de efectos y, por tanto, a una distorsión en la estimación del efecto que se quiere estudiar. Al igual que los sesgos de selección e información, el sesgo de confusión puede aumentar o disminuir el efecto, llegando incluso a anularlo o a invertir su sentido. Por tanto, también afecta a la validez del estudio. Pero a diferencia de lo que ocurre con los sesgos de selección e información, el sesgo de confusión se puede cuantificar y, sobre todo, siendo este el gran hecho diferencial, se puede controlar en el análisis. Es decir, si no se ha podido prevenir en el diseño, el sesgo de confusión se puede controlar en la fase de análisis.

2.5. Sesgo de selección

El sesgo de selección es una distorsión del efecto medido producido por una selección incorrecta de las personas que forman los grupos del estudio. Por ello, la relación entre exposición y enfermedad es diferente para los que participan y para aquellos elegibles, pero que no participan

en el estudio. Esto lleva a una estimación del efecto en los sujetos incluidos en el estudio distinta de la que se obtendría en la población entera.

Aunque el sesgo de selección es más frecuente en los estudios de casos y controles que en los de cohortes, en cualquier estudio epidemiológico existe la posibilidad teórica de que esté presente un sesgo de selección. Este sesgo debe reconocerse mediante una reflexión razonada sobre la población base en función de la hipótesis del estudio.

Existen numerosos tipos de sesgos de selección, aunque todos tienen en común que la relación entre exposición y enfermedad es diferente para los que participan y para aquellos que siendo elegibles no participan. A continuación, sin ánimo exhaustivo y teniendo en cuenta que no se trata de una taxonomía rígida ni mutuamente excluyente, se comentan algunos de los distintos tipos de sesgos, con el objetivo de mostrar las distintas posibilidades en que un sesgo de selección puede aparecer. La denominación de cada tipo de sesgo es tema de discusión. Por ello, el interés fundamental más que en poner nombre a cada sesgo, debe centrarse en considerar la posibilidad de que esté presente el sesgo, en qué medida puede estar actuando y en valorar su sentido (positivo o negativo).

a) Voluntarios.

b) No respuesta.

c) Abandonos del estudio.

d) Trabajador sano.

e) Sesgo de Berkson.

f) Sesgo de prevalencia o sesgo de supervivencia (sesgo de Neyman).

g) Sesgo de detección.

h) Sesgo por inclusión / exclusión.

A) VOLUNTARIOS

La autoselección va a producir un sesgo de selección. Los voluntarios (autoseleccionados) suelen presentar características particulares que les convierten en un grupo diferenciado. Algunas de estas características pueden estar asociadas con la exposición, la enfermedad o con ambas.

Ejemplo: En el estudio sobre la leucemia entre los militares expuestos en la prueba atómica de Smoky, realizada en 1957 en Nevada (Estados Unidos), se consiguió estudiar al 76% de todos los expuestos. Los investigadores localizaron al 82% de los estudiados y el 18% restante contactaron por propia iniciativa (autoderivación). Hubo cuatro casos de leucemia en el 82% de las personas reclutadas por los investigadores y otros cuatro casos en el 18% que se habían autoderivado.

Si las razones de la autoderivación están asociadas con el fenómeno que se estudia, bien porque su salud sea mala o porque estén particularmente expuestos, la validez del estudio se va a ver afectada.

B) NO RESPUESTA

El problema de los que no responden es muy frecuente. La ausencia puede deberse a que no se haya podido encontrar al sujeto seleccionado o a su negativa a contestar. En ambos casos, los que no responden pueden constituir un grupo muy diferente a la población estudiada. Puede que no se encuentre a una persona porque haya fallecido o se encuentre enferma; y si se niega a contestar puede que lo haga por estar afectada por el fenómeno estudiado. En definitiva, se puede producir un sesgo por las diferencias entre los participantes y los no participantes.

El sesgo de no respuesta se puede contrarrestar o disminuir realizando varios intentos de captar a los sujetos elegidos para participar, reduciendo al mínimo el número de no respuestas. En último extremo se recogerá información de quienes no contestan, para describir sus características principales (edad, sexo, etc.) y compararlas con las de los que sí contestan. Esto va a permitir valorar la comparabilidad de los sujetos participantes y no participantes

C) ABANDONOS DEL ESTUDIO

Las pérdidas, ausencias o abandonos del estudio constituyen un caso particular de autoselección. Se trata de personas que inicialmente entraron dentro de un estudio de seguimiento, pero que posteriormente, cuando se realiza el seguimiento de la cohorte el propio sujeto decide no continuar, ha fallecido por un problema no relacionado con la investigación, ha cambiado de domicilio, o no se le pueda localizar por cualquier otro motivo. Este fenómeno es más frecuente cuanto más largo es el periodo de observación y el problema será tanto mayor cuantas más pérdidas se produzcan. La primera consecuencia de la disminución del tamaño del estudio será la disminución de su potencia o poder (complementario del error beta o probabilidad de detectar como estadísticamente significativo / rechazar la hipótesis nula cuando realmente existe una asociación en la población). La validez interna del estudio no se verá afectada si las pérdidas ocurren por igual en expuestos y no expuestos, ya que la tasa de enfermedad de las personas perdidas será la misma que la de las personas del grupo del que procedían. Sin embargo, cuando los abandonos del estudio no ocurren al azar en los dos grupos de seguimiento (expuestos y no expuestos) y se relacionan con la respuesta, es decir, las pérdidas están relacionadas con la exposición y con la enfermedad, la tasa de enfermedad en las pérdidas será distinta de la que tiene la cohorte donde se encontraban y, en este caso, la validez interna se verá afectada.

Las pérdidas durante el seguimiento se pueden valorar fácilmente respecto a la exposición, ya que se conoce la distribución de los factores de riesgo en las pérdidas y se puede comparar orientativamente la frecuencia de las distintas variables entre los sujetos que permanecen en el estudio y las pérdidas. Sin embargo, es más difícil valorar cómo se van a comportar las pérdidas frente a la enfermedad, por la dificultad de conocer la tasa de enfermedad en un grupo de personas que ya no está bajo vigilancia. Por ello, al igual que en la no respuesta, hay que esforzarse por reducir al mínimo el número de ausencias, intentando encontrarlos nuevamente.

El sesgo de las pérdidas es característico de los estudios de seguimiento.

D) TRABAJADOR SANO

Incluso antes de que los sujetos sean seleccionados para el estudio se puede producir una autoselección. Un ejemplo es el denominado efecto del trabajador sano. Los trabajadores suelen tener mejor salud que la población general, ya que se habrían autoseleccionado aquellas personas con mejor salud para obtener un puesto de trabajo y poder mantenerlo, mientras que la población general también incluye enfermos incapaces de desempeñar un trabajo. Este efecto es tanto mayor cuanto más difíciles sean las condiciones de trabajo y, por tanto, requiera más capacidades singulares.

Se da siempre y cuando el estado de salud previo a entrar en una empresa pueda estar relacionado con la enfermedad. Se podría evitar mediante una definición más restringida de la población base, por ejemplo, si se toma como población no expuesta a trabajadores de la misma empresa o de otra de características similares (restricción de la población base).

Este sesgo habitualmente va a producir una infraestimación del efecto estimado.

E) SESGO DE BERKSON

Descrito por Berkson en 1946, el llamado sesgo o paradoja (término preferido por el autor) de Berkson se puede producir cuando los sujetos del estudio se obtienen del hospital (casos y controles hospitalarios). Las personas hospitalizadas pueden diferir de manera sistemática de la población general, de la que se pretende sean representivas, en muchos aspectos, y particularmente en cuanto a la exposición a los factores de riesgo estudiados, debido a diversos factores que influyen en la probabilidad de hospitalización.

El sesgo de Berkson se produce cuando la probabilidad de hospitalización de los casos y de los controles, o de los expuestos y no expuestos es distinta. En este caso, si no se consideran las distintas probabilidades de hospitalización, se pueden producir asociaciones espurias entre una exposición y una enfermedad. No se va a producir el sesgo de Berkson cuando la exposición que se estudia no influye en la hospitalización, ni cuando las poblaciones de las que se obtienen los casos y los controles sean mutuamente excluyentes (por ejemplo cuando los casos son hospitalarios y los controles son comunitarios).

El sesgo de Berkson es un sesgo de selección característico de los estudios de casos y controles.

Ejemplo: En el clásico estudio de casos y controles de Doll y Hill (1950) sobre la relación del consumo de tabaco y el cáncer de pulmón se utilizaron controles hospitalarios. El consumo de tabaco produce numerosas patologías, algunas de ellas de elevada prevalencia como la enfermedad pulmonar obstructiva crónica, que pueden precisar ingreso hospitalario. Esto hace que los controles no sean representativos en cuanto a la exposición al factor de riesgo (consumo de tabaco) de la población de donde proceden los casos, incumpliéndose la condición central para que un estudio de casos y controles sea válido, que los controles sean seleccionados independientemente de su estatus de exposición. En este estudio, dado que la relación entre el consumo de tabaco y el cáncer de pulmón es muy fuerte, a pesar de la presencia del sesgo de Berkson, que subestima el efecto, se pudo poner de manifiesto una asociación.

F) SESGO DE PREVALENCIA O SESGO DE SUPERVIVENCIA (SESGO DE NEYMAN)

Descrito por Neyman en 1955, se produce cuando se eligen casos prevalentes en lugar de casos incidentes, si la exposición es un factor pronóstico de la enfermedad. Los casos prevalentes no representan la historia natural de la enfermedad, los casos muy graves (mueren) y los leves (sanan pronto) tienen menor probabilidad de ser seleccionados. Por tanto, los casos prevalentes son supervivientes de la enfermedad y no tienen porqué tener una distribución similar de los distintos factores de riesgo de la enfermedad con respecto al total de casos.

El sesgo de prevalencia se puede evitar con la utilización exclusiva de casos incidentes en los estudios de casos y controles.

G) SESGO DE DETECCIÓN

Se produce cuando el estatus de exposición influye en la selección. En las personas que por estar expuestas o por presentar una clínica relacionada con la exposición, tienen una vigilancia mayor, se va a facilitar la detección de la enfermedad. De esta forma se va a producir un sesgo de selección.

Ejemplo: En los estudios de casos y controles sobre la administración de estrógenos en mujeres y cáncer de endometrio, dado que los estrógenos pueden producir metrorragias esto llevaría a las mujeres a solicitar atención sanitaria. Como la metrorragia es un signo del cáncer de endometrio, se incrementaría su búsqueda en el grupo que toma estrógenos, detectándose más casos de enfermedad, no tanto porque los estrógenos los produjeran sino

porque facilitarían su diagnóstico. De esta forma se produciría un sesgo de selección, ya que los casos del estudio estarían más expuestos y se sobreestimaría el efecto. Para paliar el problema se propuso que los controles fueran mujeres con enfermedades ginecológicas benignas que estuvieran también sujetas al sesgo de detección.

H) SESGO POR INCLUSIÓN / EXCLUSIÓN

Este sesgo puede darse cuando se incorporan o excluyen sistemáticamente otras enfermedades que están relacionadas con la exposición que se estudia. Este sesgo se produce por la multiefectividad de la exposición (un factor de riesgo se asocia a varias enfermedades).

El sesgo por inclusión / exclusión es un sesgo de selección característico de los estudios de casos y controles.

Ejemplo: Los anticonceptivos orales se asocian a numerosos procesos (cáncer de endometrio, cáncer de ovario, mastopatías benignas, etc.). En un estudio de casos y controles sobre los anticonceptivos orales y el cáncer de mama si se seleccionan los controles entre las mujeres atendidas en una consulta de ginecología que no tengan cáncer de mama, se producirá un *sesgo de inclusión* que infraestimará el efecto, ya que los anticonceptivos orales se asocian con numerosas patologías. Sin embargo, si del grupo control se excluyen todos los procesos relacionados con los anticonceptivos orales, se producirá un *sesgo de exclusión* si esos procesos no se eliminan también del grupo de los casos. Es decir, si se excluyen todos los casos de cáncer de mama que padezcan también un proceso relacionado con los anticonceptivos orales. Si no se procede de esta forma, habrá un sesgo de exclusión, que será proporcional a la frecuencia con que se asocia el cáncer de mama con otras patología derivadas del consumo de anticonceptivos orales.

Habrá que proceder con equilibrio para evitar el sesgo de inclusión y el de exclusión. Habitualmente, se admite mejor el sesgo por exclusión que el de inclusión, ya que el error introducido por el sesgo de exclusión es de menor magnitud.

2.6. Sesgo de información

El sesgo de información es una distorsión del efecto medido producido por un error en la obtención de la información, es decir, en la recogida de datos. Ocurre entre personas ya incluidas en el estudio, a la hora de medir las variables de interés.

El sesgo de clasificación suele ser el sesgo de información más frecuente. El sesgo de clasificación se puede producir por un:

— Error en el procedimiento de medida (problemas de calibración del instrumento, problemas de recuerdo, uso de la historia clínica, etc.).

— Error por usar variables sucedáneas (*proxy* en inglés) que se aproximan a la variable de interés. Por ejemplo, usar el peso y la talla autodeclarados para calcular el índice de masa corporal en lugar del peso y talla medidos.

— Error en la definición de las variables, algunas de ellas por su propia dificultad intrínseca como la clase social.

El sesgo de clasificación se va a producir siempre que haya errores en la clasificación de la enfermedad y/o de la exposición de las personas ya incluidas en el estudio. Los errores en la valoración de la exposición suelen ser más frecuentes que los de la enfermedad. Su magnitud va a depender de la sensibilidad y especificidad del método empleado para medir las variables, pero sus consecuencias van a diferir según si el error de clasificación es diferencial o no.

2.6.1. Clasificación errónea no diferencial

Se producirá una clasificación errónea no diferencial cuando la clasificación errónea de los sujetos en las categorías de enfermedad (o de exposición) se realiza de la misma manera en los expuestos y no expuestos (o enfermos y no enfermos), es decir, si la sensibilidad y especificidad son iguales entre los grupos a comparar. Dicho de otra manera, cuando los errores cometidos en la recogida de la información relativa a la enfermedad (o a la exposición) -en un eje- son independientes del estado de exposición (o de enfermedad) -el otro eje-.

En los *estudios de casos y controles*, el sesgo de clasificación errónea no diferencial se debe sobre todo a la medición de la exposición y produce siempre un sesgo hacia el valor nulo (infraestima el efecto). El sesgo de clasificación errónea no diferencial del estatus de enfermedad también introduce un sesgo hacia el valor nulo.

En la Tabla 4.2. se muestra un ejemplo hipotético del efecto de la clasificación errónea no diferencial de la exposición en un estudio de casos y controles, que infraestima el efecto disminuyendo la OR real de 4 a una OR sesgada de 3,5.

Tabla 4.2. Ejemplo hipotético del efecto de la clasificación errónea no diferencial de la exposición en un estudio de casos y controles.

Para casos y controles
Sensibilidad: 100%
Especificidad: 100%

	Expuestos	**No expuestos**
Casos	100	50
Controles	50	100

OR = 4

Para casos y controles
Sensibilidad: 90%
Especificidad: 100%

	Expuestos	**No expuestos**
Casos	90	60
Controles	45	105

OR = 3,5

En los *estudios de cohortes* el efecto del sesgo de clasificación errónea no diferencial depende de la medida del efecto que se estime (RA o RR) y de la dirección en que ocurra la clasificación incorrecta (sobredimensionando o, más frecuentemente, infraestimando). Si sólo se reconocen y contabilizan una proporción de los casos reales (*infracontabilización* de los casos), pero esta proporción es igual para los expuestos que para los no expuestos y, además no hay sobrecontabilización, el RA estará sesgado hacia el valor nulo (valor del RA sin sesgo multiplicado por la proporción de casos contabilizados) ($RA_{sesgado} = RA_{sin\ sesgo}$ x proporción de casos contabilizados), mientras que el RR no se verá afectado.

En la Tabla 4.3 se muestran los valores de las medidas de efecto (RA y RR) de un ejemplo hipotético de un estudio de cohortes con clasificación correcta, es decir, sin sesgo de clasificación. Por su parte, en la Tabla 4.4. se muestra un ejemplo hipotético del efecto de la clasificación errónea no diferencial en ese estudio de cohortes, infracontabilizando un 10 %

de los casos expuestos y no expuestos. Se aprecia, en comparación con los valores de la tabla 4.3 que el RA está sesgado hacia el valor nulo, lo infraestima disminuyendo el RA real de 0,2 a un RA sesgado de 0,18 (valor del RA sin sesgo - 0,2 - multiplicado por la proporción de casos contabilizados - 0,90-)(0,2 x 0,90 = 0,18). Mientras que el RR no está afectado, RR de 1,5.

Si en un estudio de cohortes se *sobrecontabilizan* los casos, añadiendo como casos sujetos adicionales que no padecen la enfermedad, pero estos sujetos se añaden con proporcionalidad a los tamaños totales de los denominadores de expuestos y no expuestos, el RA no se verá afectado, mientras que el RR estará sesgado hacia el valor nulo.

Tabla 4.3. Ejemplo hipotético de un estudio de cohortes con clasificación correcta (sin sesgo de clasificación).

	Expuestos	**No expuestos**	
Casos	60	40	100
No casos	40	60	100
	100	100	200

$$Ie = \frac{a}{a+c} = \frac{60}{100} = 0,6$$

$$Io = \frac{b}{b+d} = \frac{40}{100} = 0,4$$

$$RA = Ie - Io = 0,6 - 0,4 = 0,2$$

$$RR = \frac{Ie}{Io} = \frac{0,6}{0,4} = 1,5$$

Tabla 4.4. Ejemplo hipotético de un estudio de cohortes con clasificación errónea no diferencial (infracontabilizando un 10% de los casos expuestos y no expuestos).

	Expuestos	No expuestos	
Casos	60 - 6 = 54	40 - 4 = 36	90
No casos	46	64	110
	100	100	200

$$Ie = \frac{54}{100} = 0,54$$

$$Io = \frac{36}{100} = 0,36$$

$$RA = Ie - Io = 0,54 - 0,36 = 0,18$$

RA sesgado hacia el valor nulo

$$RA = 0,2 \text{ x } 0,9 = 0,18$$

RA = valor sin sesgo (0,2) x proporción de casos contabilizados (0,9) = 0,18

$$RR = \frac{Ie}{Io} = \frac{0,54}{0,36} = 1,5$$

RR no se afecta

En la Tabla 4.5. se muestra un ejemplo hipotético del efecto de la clasificación errónea no diferencial en un estudio de cohortes, sobrecontabilizando los casos un 10 % sobre el tamaño total de los denominadores de expuestos y no expuestos. Se aprecia, en comparación con los valores de la tabla 4.3, que no tienen sesgo, que el RA no está afectado, RA de 0,2. Mientras que el RR está sesgado hacia el valor nulo, lo infraestima disminuyendo el RR real de 1,5 a un RR sesgado de 1,4.

Tabla 4.5. Ejemplo hipotético de un estudio de cohortes con clasificación errónea no diferencial (sobrecontabilizando los casos un 10% sobre el tamaño total de los denominadores de expuestos y de no expuestos).

	Expuestos	**No expuestos**	
Casos	60 + 10 = 70	40 + 10 = 50	120
No casos	30	50	80
	100	100	200

$$Ie = \frac{70}{100} = 0,7$$

$$Io = \frac{50}{100} = 0,5$$

$$RA = Ie - Io = 0,7 - 0,5 = 0,2$$

RA no se afecta

$$RR = \frac{Ie}{Io} = \frac{0,7}{0,5} = 1,4$$

RR sesgado hacia el valor nulo

2.6.2. Clasificación errónea diferencial

Se producirá una clasificación errónea diferencial cuando la clasificación errónea de los sujetos en las categorías de enfermedad (o de exposición) se realiza de manera diferente según el sujeto esté expuesto o no expuesto (o enfermo o no enfermo), es decir, si la sensibilidad y especificidad difieren entre los grupos a comparar. Dicho de otra manera, cuando los errores cometidos en la recogida de la información relativa a la enfermedad (o a la exposición) -en un eje- no son independientes del estado de exposición (o de enfermedad) -el otro eje-.

El sesgo de clasificación errónea diferencial puede producir un sesgo en cualquier dirección aumentando o disminuyendo la asociación real.

En la Tabla 4.6. se muestra un ejemplo hipotético del efecto de la clasificación errónea diferencial de la exposición en un estudio de casos y controles, que bien puede sobreestimar el

efecto aumentando la OR real de 4 a una OR sesgada de 4,7 o infraestimarlo disminuyendo la OR de 4 a una OR sesgada de 3.

Tabla 4.6. Ejemplo hipotético del efecto de la clasificación errónea diferencial de la exposición en un estudio de casos y controles.

Para casos y controles
Sensibilidad: 100%
Especificidad: 100%

	Expuestos	No expuestos
Casos	100	50
Controles	50	100

OR = 4

Casos:
Sensibilidad: 100%
Especificidad: 100%

Controles:
Sensibilidad: 90%
Especificidad: 100%

	Expuestos	No expuestos
Casos	100	50
Controles	45	105

OR = 4,7

Casos:
Sensibilidad: 90%
Especificidad: 100%

Controles:
Sensibilidad: 100%
Especificidad: 100%

	Expuestos	No expuestos
Casos	90	60
Controles	50	100

OR = 3

La clasificación errónea no diferencial se considera una amenaza para la validez menor que la diferencial, ya que va siempre en una dirección predecible (hacia la condición nula) y es más aceptable subestimar los efectos que sobrestimarlos.

La clasificación errónea no diferencial, que actúa prácticamente en todo estudio epidemiológico, es crucial en estudios que indican poco o ningún efecto. Este sesgo explicaría muchas discrepancias entre estudios epidemiológicos.

2.6.3. Tipos de sesgos de clasificación

Existen numerosos tipos de sesgos de clasificación. A continuación, sin ánimo exhaustivo y teniendo en cuenta que no se trata de una taxonomía rígida ni mutuamente excluyente, se comentan algunos de los distintos tipos de sesgos, con el objetivo de mostrar las distintas

posibilidades en que un sesgo de clasificación puede aparecer. La denominación de cada tipo de sesgo es tema de discusión. Por ello, el interés fundamental más que en poner nombre a cada sesgo, debe centrarse en considerar la posibilidad de que esté presente el sesgo, en qué medida puede estar actuando y en valorar su sentido (positivo o negativo).

a) Sospecha de diagnóstico o sesgo de diagnóstico.

b) Sospecha de exposición.

c) Sesgo de memoria o de recuerdo.

d) Sesgo de codificación y registro de datos.

e) Sesgo de realización de mediciones o diagnósticos.

f) Sesgo protopático o de latencia de enfermedad.

g) Sesgo por el efecto vigilancia o Hawthorne.

f) Desenmascaramiento.

A) SOSPECHA DE DIAGNÓSTICO O SESGO DE DIAGNÓSTICO

El sesgo de sospecha de diagnóstico se produce cuando se hace un mayor esfuerzo diagnóstico para detectar una enfermedad en el grupo expuesto que en el grupo no expuesto. Esto va a ocasionar una clasificación errónea diferencial, que va a sobreestimar el efecto.

El sesgo de sospecha de diagnóstico es característico de los estudios de seguimiento.

Cabe señalar las similitudes entre este sesgo de clasificación errónea diferencial y el sesgo de selección. Sin embargo, como ocurre entre sujetos ya incluidos en el estudio, se considera de clasificación.

El sesgo de diagnóstico se puede evitar asegurando un seguimiento comparable.

Ejemplo: En un estudio de seguimiento para comparar la incidencia de enfisema en fumadores y no fumadores, si los fumadores tienen una mayor atención médica que los no fumadores, se diagnosticará con más frecuencia el enfisema en los fumadores por el mero hecho de hacer un mayor esfuerzo diagnóstico para detectar una enfermedad en el grupo expuesto que en el grupo no expuesto. De esta forma se va a producir un subdiagnóstico más frecuente en no fumadores. Esto va a ocasionar una clasificación errónea diferencial, que va a sobreestimar el efecto.

B) SOSPECHA DE EXPOSICIÓN

El sesgo de sospecha de exposición se produce cuando se hace un mayor esfuerzo en buscar en los casos la exposición a factores de riesgo conocidos de la enfermedad. Esto va a ocasionar una clasificación errónea diferencial, que va a sobreestimar el efecto.

El sesgo de sospecha de exposición es característico de los estudios de casos y controles.

C) SESGO DE MEMORIA O DE RECUERDO

El sesgo de memoria, de recuerdo o anamnésico se puede producir siempre que se utilice una pregunta anamnésica, es decir, cuando la exposición dependa de la memoria, ya que puede haber un recuerdo inexacto de la exposición. Si el error de recuerdo es similar en los casos y en los controles, se producirá una clasificación errónea no diferencial. Sin embargo, es frecuente que los casos recuerden mejor la exposición que los controles, por haber estado "rumiando" sobre las posibles causas de su enfermedad. Esto va a ocasionar una clasificación errónea diferencial, que va a sobreestimar el efecto.

A veces, puede ocurrir a la inversa, el enfermo puede mostrar tendencia (inconscientemente o no) a minusvalorar una exposición si se siente responsable de su enfermedad. Un ejemplo

clásico de ello lo constituyen los bebedores de alcohol. En este caso se va a producir una clasificación errónea diferencial, que va a infraestimar el efecto.

Otro elemento que afecta al sesgo de memoria es el tiempo transcurrido entre la exposición y el recuerdo. Por ello, si el tiempo medio transcurrido desde que se produjo la exposición fuese distinto en los casos que en los controles, se podría producir un sesgo de memoria.

El sesgo de memoria es característico de los estudios de casos y controles, aunque también puede darse en los estudios de seguimiento, al inicio del estudio a la hora de categorizar a los participantes en función de su nivel de exposición.

Este sesgo se puede evitar o aminorar con el doble ciego (ni los participantes ni los encuestadores conocen la asociación que se investiga), la comprobación de las respuestas con datos más objetivos o el uso de controles hospitalarios.

Ejemplo: En un estudio de casos y controles sobre el consumo de fármacos durante el embarazo y la aparición de malformaciones congénitas, las madres de niños malformados van a recordar su exposición de forma más completa que las madres que tuvieron un hijo sano.

D) SESGO DE CODIFICACIÓN Y REGISTRO DE DATOS

También puede producirse un sesgo en la codificación o registro de los datos. Puede ser frecuente el subregistro de datos sobre factores de riesgo en las historias clínicas, por ejemplo del consumo de tabaco.

Produce un sesgo de clasificación errónea no diferencial, que infraestimará el efecto.

E) SESGO DE REALIZACIÓN DE MEDICIONES O DIAGNÓSTICOS

El sesgo de realización de mediciones o diagnósticos es debido a errores en los aparatos de medida (falta de calibración) o en la realización de las determinaciones. Esto va a ocasionar una clasificación errónea no diferencial, que va a infraestimar el efecto.

También puede ocurrir que se tienda a clasificar más como enfermos a las situaciones límite cuando están expuestos a un posible factor de riesgo de la enfermedad estudiada. Produce una clasificación errónea diferencial, que sobreestimará el efecto.

F) SESGO PROTOPÁTICO O DE LATENCIA DE ENFERMEDAD

El sesgo protopático o de latencia de la enfermedad se puede producir cuando las manifestaciones precoces de la enfermedad, que no ha sido aún detectada, condicionan un cambio en la exposición de los casos. Este sesgo puede hacer que los factores de riesgo aparezcan como factores de protección.

Algunos autores consideran el sesgo protopático como un sesgo de confusión o de selección, pero como es controlable mediante una historia cronológica correcta, se incluye como sesgo de clasificación errónea diferencial.

El sesgo protopático es característico de los estudios de casos y controles.

Ejemplo: En un estudio entre consumo de alcohol y colelitiasis se podría observar un efecto protector el consumo de alcohol debido al sesgo protopático. Las personas con colelitiasis, antes de que les sea diagnosticada, pueden tender a beber menos alcohol que la población general por las molestias que sienten.

Un caso similar podría darse en un estudio sobre la relación del consumo de salicilatos y la úlcera grastroduodenal.

G) SESGO POR EL EFECTO VIGILANCIA O HAWTHORNE

El sesgo por el efecto vigilancia o Hawthorne se produce por la modificación de la conducta habitual de una persona cuando se siente observada. Recibe este nombre por las investigaciones realizadas en la central eléctrica Western Electric's Hawthorne de Chicago de 1927 a 1937 sobre la satisfacción de los trabajadores y su productividad, que tuvieron gran influencia en la psicología industrial. En este estudio, la manipulación de las condiciones ambientales de los trabajadores aumentaban el rendimiento hasta que se extinguía la novedad de sentirse parte de un estudio.

Este sesgo puede darse, por ejemplo, en los pacientes que participan en un ensayo clínico. Se puede evitar si se incluye un grupo placebo o control, en el que también se dé el efecto vigilancia y, por tanto, la comparación no se verá afectada por este sesgo.

H) DESENMASCARAMIENTO

Se produce cuando un posible factor de riesgo de una enfermedad produce un efecto secundario semejante a un signo de la enfermedad. Este hecho, puede hacer que se intensifique la búsqueda de la enfermedad en el grupo con el factor de riesgo y, por ello, detectar más casos de enfermedad en los expuestos. Esto produce una clasificación errónea diferencial, que sobrestimará el efecto.

Ejemplo: En un estudio sobre la administración de estrógenos en mujeres posmenopáusicas y cáncer de endometrio, dado que los estrógenos pueden producir metrorragias, que son un signo del cáncer de endometrio, puede que se incremente la búsqueda del cáncer de endometrio en el grupo con toma estrógenos y, por ello, detectar más casos de enfermedad. Esto produce una clasificación errónea diferencial, que sobrestimará el efecto.

2.7. Sesgo ecológico o falacia ecológica

El sesgo ecológico (o sesgo de agregación) es un sesgo de información que se produce cuando la estimación del efecto ecológico esperado no representa necesariamente el efecto que existe a nivel individual. El sesgo ecológico constituye la principal limitación del análisis ecológico para hacer inferencias causales. Fue descrito por primera vez por Robins en 1950 en el campo de la sociología.

Cabe recordar la gráfica definición de la falacia ecológica dada por Michael, para quien "la falacia ecológica es un astuto y sigiloso animal de dos caras, que está al acecho en las esquinas oscuras de los estudios ecológicos, y engaña a los lectores incautos haciéndoles aceptar conclusiones injustificadas, ya que presenta conclusiones que pueden ser válidas o pueden ser falsas, dependiendo de la boca por la que hable".

Ejemplo: El sociólogo francés Emile Durkheim (1858-1917) encontró una relación entre la proporción de protestantes y la tasa de suicidio en las provincias de Prusia entre 1883 y 1890. Este hallazgo no puede interpretarse como que hay una relación entre ser protestante y suicidarse, ya que podría ser que en las regiones de mayoría protestante fuesen los católicos los que se suicidasen.

El sesgo ecológico será expuesto detalladamente en el Capítulo 9, que trata de los Estudios Ecológicos.

3. SESGO DE CONFUSIÓN

La confusión es un concepto clave en epidemiología. Incluso, podría llegar a entenderse la epidemiología como la ciencia que se encarga del control de los factores de confusión.

3.1. Concepto de confusión

Nuevamente, podemos tener una primera aproximación al concepto de confusión a través del diccionario de la Real Academia Española, que define el verbo confundir como mezclar, fundir cosas diversas, de manera que no puedan reconocerse o distinguirse. En su tercera acepción lo define como equivocar, tomar una cosa por otra. Esta definición proporciona la clave del fenómeno, hay una mezcla que conduce a una equivocación.

En términos epidemiológicos, se puede definir el sesgo de confusión (*confounding* en inglés) como la distorsión del efecto estimado de la exposición sobre la enfermedad debido a la interposición de un efecto nuevo producido por un factor extraño (factor de confusión, variable confusora o confusor), que no se quiere estudiar, y que origina una mezcla de efectos.

Ejemplo: Supóngase que se quiere estudiar el efecto de la exposición al consumo de café sobre el cáncer de páncreas y que esta relación estuviera afectada por la mezcla con el efecto de la exposición al consumo de tabaco. Teniendo en cuenta que en este estudio sólo se quiere analizar el efecto del consumo de café, se habría introducido un factor extraño o de confusión (consumo de tabaco), que no se quiere estudiar y que distorsiona el efecto.

3.2. Tipos de efectos de la confusión

Siempre hay que tener presente que la confusión es un sesgo, en este caso producido por la presencia de una variable externa, denominada factor de confusión, que da lugar a una mezcla de efectos y, por tanto, a una distorsión en la estimación del efecto que se quiere estudiar. El sesgo de confusión puede aumentar o disminuir el efecto, llegando incluso a anularlo o a invertir su sentido. Se denomina *confusión positiva* cuando el parámetro epidemiológico presenta un valor superior al que realmente tiene (sobreestima o exagera el efecto), es decir, lo aleja de la hipótesis nula (RR=1). Se denomina *confusión negativa* cuando el parámetro presenta un valor inferior al real (infraestima o atenúa el efecto), es decir, lo aproxima a la hipótesis nula. Se denomina paradoja de Simpson cuando la presencia de una variable de confusión cambia el sentido de una asociación (*confusión cualitativa*). De esta forma, la confusión sería capaz de hacer que un factor de riesgo apareciera como factor de protección y al revés. Aunque como señala Rothman, este fenómeno no es realmente una paradoja (hecho o dicho aparentemente contrario a la lógica), ya que no se contravienen ni la lógica ni ninguna de sus premisas, sino que es la consecuencia de no reconocer la presencia de variables de confusión.

La confusión afecta a la validez del estudio. Pero a diferencia de lo que ocurre con los sesgos de selección e información, el sesgo de confusión se puede cuantificar y, sobre todo, siendo este el gran hecho diferencial, se puede controlar en la fase de análisis. Es decir, si no se ha podido evitar en el diseño, el sesgo de confusión se puede controlar en el análisis.

3.3. Requisitos

Una vez que se ha definido el concepto de sesgo de confusión y se han expuesto sus características principales, hay que señalar que para que una variable o factor extraño se pueda considerar como un factor de confusión debe cumplir tres requisitos:

a) Estar asociado con la exposición.

b) Estar asociado con la enfermedad.

c) No ser un paso intermedio en la secuencia causal entre exposición y enfermedad.

A continuación se van a analizar con más detalle estos tres requisitos o características que debe reunir una variable extraña para que se la pueda denominar factor de confusión.

A) EL FACTOR DE CONFUSIÓN DEBE ESTAR ASOCIADO CON LA EXPOSICIÓN.

El factor de confusión debe distribuirse de forma diferente en los expuestos y los no expuestos. La asociación entre el potencial factor de confusión y la exposición no debe ser secundaria a una asociación entre el factor de confusión y la enfermedad. Por ello, se estudiará la asociación del factor de confusión con la exposición en los no enfermos.

Ejemplo: En el estudio sobre el consumo de café y cáncer de páncreas, donde el consumo de tabaco puede ser un hipotético factor de confusión. El consumo de tabaco debe estar asociado con el consumo de café en los no casos si fuera un estudio de cohortes (RR distinto de 1) o en los controles si se tratara de un estudio de casos y controles (OR distinto de 1).

B) EL FACTOR DE CONFUSIÓN DEBE ESTAR ASOCIADO CON LA ENFERMEDAD.

Para que un factor extraño confunda debe producir un efecto, debe ser predictivo de la ocurrencia de enfermedad, es decir, debe ser un factor de riesgo de la enfermedad estudiada, pero no es necesario que sea causal (por ejemplo la edad). La asociación entre el potencial factor de confusión y la enfermedad no debe ser secundaria a una asociación entre el factor de confusión y la exposición. Por ello, se estudiará la asociación del factor de confusión con la enfermedad en los no expuestos.

Ejemplo: En el estudio sobre el consumo de café y cáncer de páncreas, donde el consumo de tabaco puede ser un hipotético factor de confusión. El consumo de tabaco debe estar asociado con el cáncer de páncreas en los que no consumen café (RR u OR distinto de 1). Es decir, en los no consumidores de café, el consumo de tabaco es un factor de riesgo del cáncer de páncreas.

C) EL FACTOR DE CONFUSIÓN NO DEBE SER UN PASO INTERMEDIO EN LA SECUENCIA CAUSAL ENTRE EXPOSICIÓN Y ENFERMEDAD.

El hipotético factor de confusión no debe ser un paso intermedio en la cadena causal entre exposición y enfermedad. Por este motivo se la llama variable extraña. Esto implica que la capacidad predictiva del riesgo de enfermedad contemple un mecanismo diferente del que se está investigando. Todo factor que sea un paso intermedio en la secuencia causal no podrá ser un factor de confusión. Determinar este criterio requiere un conocimiento previo de las causas de la relación estudiada, basándose en el conocimiento existente sobre la relación estudiada y el juicio experto.

Ejemplo: En el estudio sobre el consumo de café y cáncer de páncreas, donde el consumo de tabaco puede ser un hipotético factor de confusión. Conforme al conocimiento actualmente disponible, el consumo de tabaco no es un paso intermedio en la vía causal entre el consumo de café y el cáncer de páncreas.

Para que a una variable extraña se la considere como factor de confusión debe cumplir los tres requisitos. Si la variable extraña sólo está asociada a la exposición o sólo está asociada a la enfermedad, no será un factor de confusión. Los dos primeros criterios se podrán comprobar con los datos del propio estudio. Sin embargo, el tercer criterio requiere información externa a los datos.

3.4. Cuantificación

Una de las peculiaridades del sesgo de confusión, que lo diferencia de los sesgos de selección e información, es que se puede cuantificar. Por ello, lo importante no es tanto determinar su presencia o ausencia como el poder cuantificar su efecto.

La evaluación de la magnitud de la confusión se realizará mediante la comparación directa de la estimación bruta del efecto y la estimación no confundida o ajustada. Supóngase un estudio de casos y controles en el que, a partir de los datos brutos, primero se calcula el OR_{bruto} de la relación estudiada. A continuación, los datos se dividen (estratifican) en dos categorías de un hipotético factor de confusión (presente o ausente) y se calcula el OR_e de la relación estudiada en cada estrato y el $OR_{ajustado}$. De esta forma se podría cuantificar el efecto de la confusión. Si el OR_{bruto} es mayor que el $OR_{ajustado}$, se calculará el porcentaje de exceso de riesgo explicado (ERE) por la variable ajustada, mediante la fórmula:

$$\% \, ERE = \frac{ORbruto - ORajustado}{ORbruto - 1} \, x \, 100$$

El mecanismo representado por la variable de ajuste (factor de confusión) se puede considerar responsable del porcentaje de exceso de riesgo explicado (ERE) de la asociación estimada, o lo que es lo mismo, la distorsión del factor de confusión es responsable del porcentaje de exceso de riesgo explicado (ERE) de la asociación estimada.

Cuando el OR_{bruto} es menor que el $OR_{ajustado}$, se calculará el porcentaje de defecto de riesgo explicado (DRE) por la variable ajustada, mediante la fórmula:

$$\% \, DRE = \frac{ORajustado - ORbruto}{ORbruto - 1} \, x \, 100$$

El mecanismo representado por la variable de ajuste (factor de confusión) se puede considerar responsable del porcentaje de defecto de riesgo explicado (DRE) de la asociación estimada, o lo que es lo mismo, la distorsión del factor de confusión es responsable del porcentaje de defecto de riesgo explicado (DRE) de la asociación estimada.

Ejemplo: En el estudio sobre el consumo de café y cáncer de páncreas, donde el consumo de tabaco cumple los criterios como para ser considerado un factor de confusión y se ha descartado previamente la presencia de interacción. El efecto estimado a partir de los datos brutos del consumo de café sobre el cáncer de páncreas proporciona un OR_{bruto} de 2,5. Sin embargo, cuando se estratifica en dos categorías (fumadores y no fumadores), el efecto estimado del consumo de café sobre el cáncer de páncreas tanto en los fumadores como en los fumadores proporciona un OR_e, y consecuentemente el $OR_{ajustado}$ es de 2. En este caso, al ser el OR_{bruto} es mayor que el $OR_{ajustado}$, se calcula el porcentaje de exceso de riesgo explicado (ERE) por la variable ajustada, mediante la fórmula:

$$\% \, ERE = \frac{ORbruto - ORajustado}{ORbruto - 1} \, x \, 100 = \frac{2,5 - 2}{2,5 - 1} \, x \, 100 = 33,3\%$$

El mecanismo representado por la variable de ajuste (factor de confusión, en este ejemplo el consumo de tabaco) se puede considerar responsable del 33,3% de la asociación estimada, es decir, la distorsión del consumo de tabaco es responsable del 33,3% de la asociación estimada

Como se acaba de ver en las fórmulas de cálculo anteriores, la comparación del OR_{bruto} se hace con el $OR_{ajustado}$. Este $OR_{ajustado}$ es un estimador global de la relación estudiada, que ya no estaría distorsionado por el factor de confusión, sino ajustado por él. El cálculo del $OR_{ajustado}$ puede hacerse por un método exacto de máxima probabilidad o verosimilitud (*maximum likelihood* en inglés) o por el método aproximado propuesto por Mantel y Haenszel (1959), que

proporciona un resultado casi idéntico al estimado de máxima probabilidad pero que es enormemente más sencillo de calcular.

Aunque la comparación entre la estimación bruta y la ajustada no se realiza mediante tests estadísticos, sino mediante la comparación directa, con frecuencia se considera que la confusión es suficientemente importante como para tener que controlarla cuando el OR bruto difiere del ajustado en más de 10 %.

En definitiva, como señala Rothman, la decisión sobre si un factor es de confusión o no debe tomarse basándose en la mejor información disponible, sazonado con un poquito de juicio experto. La confusión nunca debe valorarse mediante pruebas estadísticas.

3.5. Procedimiento para detectar y evaluar la confusión

Para poder detectar la presencia de un factor de confusión, lo primero que habría que hacer es sospechar su presencia mediante la pregunta: ¿existe alguna explicación plausible alternativa a la presunta asociación causa-efecto objeto de estudio?

En segundo lugar, se procedería a la identificación del hipotético factor de confusión mediante la recogida de los datos disponibles sobre todos los factores de riesgo conocidos de la enfermedad estudiada. Como el factor de confusión tiene que ser un factor de riesgo de la enfermedad, la forma más operativa de empezar la búsqueda de un factor de confusión puede ser a través del estudio del resto de los factores de riesgo que no son objeto de la investigación.

En tercer lugar, se comprobará que el hipotético factor de confusión cumple los tres requisitos de condicionalidad.

Finalmente, se evaluará la confusión mediante la comparación directa de la estimación bruta del efecto y la estimación de este no confundida. Cuando se considere necesario, se aplicarán los procedimientos analíticos disponibles (análisis estratificado y análisis multivariable), que se verán a continuación, para controlar el efecto de la confusión sobre la relación estudiada.

3.6. Control de la confusión

Se dispone de distintos procedimientos que permiten prevenir la presencia de la confusión en la fase de diseño (aleatorización, restricción o especificación y emparejamiento) o controlarla en la fase de análisis (estratificación y análisis multivariable).

3.6.1. En el diseño

Se emplean tres métodos para prevenir la presencia de la confusión en la fase de diseño: la aleatorización, la restricción o especificación y el emparejamiento. La aleatorización solo se puede aplicar en los estudios experimentales. La restricción y el emparejamiento se pueden utilizar en los estudios experimentales y en los observacionales. Estos dos procedimientos implican una alteración en la selección de los sujetos estudiados para asegurar que sólo se comparan grupos similares de las potenciales variables de confusión. Esta es una decisión que se toma a priori y que es irreversible. Por ello, aunque estos dos métodos son efectivos en la prevención de la confusión, se deben usar con sumo cuidado ya que van limitar los resultados obtenidos de forma irreversible.

A) ALEATORIZACIÓN

La aleatorización o asignación aleatoria consiste en distribuir a cada participante en los grupos del estudio aleatoriamente, de modo que cada persona tenga exactamente la misma probabilidad de formar parte de un grupo u otro. Su principal ventaja es que todos los potenciales

factores de confusión conocidos y desconocidos son controlados, ya que se distribuyen de forma homogénea en todos los grupos del estudio. Para que esto sea cierto se requiere que el tamaño del estudio sea grande, siendo menos efectiva en los estudios pequeños. Además, tiene la limitación de que sólo es posible realizarla en los estudios experimentales.

B) RESTRICCIÓN O ESPECIFICACIÓN

La restricción o especificación consiste en seleccionar en el estudio únicamente a las personas que presenten un valor determinado del posible factor de confusión. Entre sus ventajas destaca que es un procedimiento eficaz, económico y cómodo, que es muy fácil de comprender. Sin embargo, proporciona una imagen parcial de la realidad, que limita la generalización de los resultados. En ocasiones puede hacer difícil obtener un tamaño del estudio apropiado.

Ejemplo: En el estudio sobre el café y el cáncer de páncreas, donde el consumo de tabaco pudiera ser un factor de confusión, se podría especificar que sólo se incluyese en el estudio a los no fumadores. De esta forma, si se observase una asociación entre el café y el cáncer de páncreas, evidentemente no podría ser debida al consumo de tabaco.

C) EMPAREJAMIENTO

El emparejamiento (*matching* en inglés) se usa principalmente en los estudios de casos y controles, pero puede emplearse en los demás tipos de diseños de estudios. En los estudios de seguimiento se utiliza pocas veces porque es demasiado costoso. En un estudio de casos y controles consistiría en seleccionar, para cada caso, un control que tuviese el mismo valor del posible factor de confusión. Esto permite comparar sólo a casos y controles que presentan el mismo nivel del posible factor de confusión. Este procedimiento puede evitar la aparición de confusión asociada a factores constitucionales (edad, sexo) o a factores que no pueden medirse ni controlarse de otra forma (emparejar hermanos para controlar factores genéticos y familiares). En algunas ocasiones puede facilitar la selección de los controles. El emparejamiento mantiene la capacidad de generalizar. Sin embargo, frecuentemente suele ser más difícil, lento, caro y menos eficiente. La decisión de emparejar debe tomarse al inicio del estudio y es irreversible. Impide estudiar el efecto de las variables emparejadas. Existe el peligro de sobreemparejamiento, es decir, emparejar por un factor que no es un factor de confusión, esto va a producir una disminución del poder del estudio.

Los estudios que utilizan el emparejamiento requieren un análisis específico, que tenga en cuenta el hecho de que los datos están emparejados.

Ejemplo: En el estudio sobre el café y el cáncer de páncreas, donde el consumo de tabaco pudiera ser un factor de confusión, podría compararse el consumo de café de un caso que fumase un paquete al día con el consumo de café de un control que fumase la misma o parecida cantidad de cigarrillos al día.

3.6.2. En el análisis

Posteriormente, en la fase de análisis se puede controlar la confusión mediante la estratificación y el análisis multivariable. Se trata de procedimientos que son más versátiles e igual de efectivos que la restricción y emparejamiento. Es importante señalar que aunque estos dos procedimientos se aplican en el análisis, ya en el diseño se tuvo que determinar qué posibles factores de confusión podrían afectar a la relación estudiada, para recoger la información sobre ellos, ya que no se podrá controlar en el análisis ninguna variable que no se haya medido previamente.

A) ESTRATIFICACIÓN

La estratificación consiste en distribuir a los sujetos en estratos de acuerdo con el nivel del posible factor de confusión, analizando por separado la relación entre la exposición y la enfermedad

en cada estrato. Este procedimiento proporciona resultados fácilmente interpretables y es flexible y reversible. Sin embargo, sólo permite controlar simultáneamente un número limitado de variables, ya que el tamaño de cada estrato se verá reducido con cada variable por la que se estratifique. Por ejemplo, para poder realizar una estratificación por cinco variables, aunque sólo tuviesen dos categorías cada una, se necesitarían 32 (2^5) estratos. Con tantos estratos, el número de personas disponibles para el análisis en cada uno puede resultar demasiado pequeño.

Además, cuando los estratos son demasiado grandes puede que no se controle adecuadamente la confusión si se produjese confusión dentro de cada estrato (confusión residual). Las variables a controlar deben haber sido medidas con antelación, es decir, en la fase de diseño se debe decidir que variables hipotéticamente podrían actuar como factores de confusión. Posteriormente, en el análisis se valorará en qué medida producen confusión.

Ejemplo: En el estudio sobre el café y el cáncer de páncreas, donde el consumo de tabaco pudiera ser un factor de confusión, se podrían considerar por separado los fumadores y los no fumadores (estratificando según el hábito de fumar). Se estudiaría la asociación entre el café y el cáncer de páncreas por una parte en fumadores y por otra en no fumadores. De esta forma, se eliminan los posibles efectos de la confusión creados por el consumo de tabaco en la asociación entre el café y el cáncer de páncreas.

B) ANÁLISIS MULTIVARIABLE

El análisis multivariable es un método analítico que permite el estudio simultáneo de dos o más variables independientes. Su principal ventaja es que permite controlar simultáneamente múltiples factores de confusión. Además, permite utilizar toda la información obtenida con las variables continuas y es flexible y reversible. Como inconvenientes pueden señalarse que en ocasiones puede proporciona resultados difíciles de interpretar. Además, el modelo puede no ser adecuado, bien por un control incompleto de los factores de confusión o por una estimación inexacta de la magnitud del efecto. También exige que las variables a controlar hayan sido medidas con antelación.

3.7. Confusión residual

Se denomina confusión residual a la confusión que persiste tras los intentos infructuosos por controlarla. Es decir, cuando el ajuste no es capaz de eliminar completamente el efecto de confusión. Las fuentes más importantes de confusión residual o ajuste incompleto son:

— Falta de ajuste por algún factor de confusión.

— Definición inadecuada de las categorías de la variable de confusión, que hace que alguna de estas categorías sea tan amplia que no se pueda asumir la ausencia de confusión dentro de cada estrato, por lo que se produce un ajuste imperfecto.

— Utilización de un sucedáneo imperfecto como variable de ajuste para controlar un factor de confusión. Es decir, cuando hay problemas en la validez por constructo (grado en que la variable que se utiliza en el ajuste representa al factor de confusión). La utilización de la educación como variable de ajuste cuando se pretende controlar el efecto como factor de confusión de la clase social puede constituir un buen ejemplo.

— Clasificación errónea de la variable de confusión.

4. INTERACCIÓN O MODIFICACIÓN DEL EFECTO

Una vez que se han expuesto los sesgos y particularmente el sesgo de confusión, es importante introducir otro concepto, como es el de interacción, que también puede producir una

modificación del efecto estimado, pero, a diferencia de los anteriores, no lo hace de forma errónea. Tanto la interacción como el sesgo de confusión, aunque son conceptualmente muy distintos, comparten la necesidad de valorar variables externas para poder interpretar correctamente el efecto estudiado.

La interacción se da cuando el efecto de la exposición sobre la enfermedad es diferente a distintos niveles de una tercera variable. Se trata de un fenómeno natural, que existe independientemente del diseño del estudio. El sesgo de confusión, por contra, es una distorsión errónea del efecto, que puede estar presente o no, dependiendo del diseño del estudio.

Recibe el nombre de *sinergismo* cuando la presencia de un factor aumenta el efecto de otro. El opuesto del sinergismo es el *antagonismo*, es decir, cuando la presencia de un factor disminuye el efecto de otro. Una interacción se considera *cuantitativa* cuando sólo cambia la magnitud de los efectos, pero no su sentido. Sin embargo, cuando una interacción cambia el sentido de la asociación se denomina *cualitativa*. Por ejemplo, cuando el efecto de la exposición sobre la enfermedad es protector en unos niveles de una tercera variable y en otros niveles es un factor de riesgo.

La interacción sí se evalúa mediante un test estadístico (test de homogeneidad de efectos-método de Woolf), junto con la valoración subjetiva de su relevancia. Téngase en cuenta que este test como todo test de significación estadística se ve afectado por el tamaño del estudio. Por ello, si el tamaño del estudio es reducido, puede que haya interacción y no se ponga de manifiesto. Sin embargo, como se ha señalado con anterioridad, el sesgo de confusión nunca debe valorarse mediante pruebas estadísticas, sino mediante la comparación directa de la estimación bruta del efecto y la estimación de este no confundido.

Aunque la interacción es un fenómeno natural, no debe ignorarse, ya que si se combinaran los datos se perdería información valiosa de posible significación biológica. La presencia de interacción impide la obtención de un estimador global del efecto, por lo que se deben considerar los efectos separadamente en cada nivel de interacción. La interacción una vez detectada se debe describir. Sin embargo, el sesgo de confusión debe evitarse en la fase de diseño del estudio o controlarse en el análisis, ya que es un error.

Ejemplo: Supóngase un estudio sobre la relación del consumo de tabaco y el cáncer boca y faringe (Rothman y Keller, 1972). El RR del consumo de tabaco con el cáncer oral en los que no beben alcohol es de 1,53 y en los que beben alcohol el RR es de 5,71. Se aprecia que la relación que se quiere estudiar (consumo de tabaco-cáncer oral) es distinta en los bebedores de alcohol que en los abstemios. Esto es debido a la presencia de una interacción entre el consumo de tabaco y el consumo de alcohol en la relación estudiada. Por ello, se deberá describir separadamente la relación consumo de tabaco-cáncer oral en los bebedores (RR=5,71) y en los abstemios (RR=1,53).

En la Tabla 4.7. se recogen las principales diferencias entre el sesgo de confusión y la interacción.

En la estrategia general de análisis de un estudio epidemiológico, antes de valorar la posible existencia de confusión se debe detectar la presencia de interacción, ya que en presencia de interacción el análisis puede concluirse presentando separadamente los estimadores específicos de cada nivel de interacción o estrato, junto con sus intervalos de confianza. En caso de que existiera también confusión, esta sólo se manifestaría en la medida cruda del efecto. Por tanto, al describir la interacción se estaría controlando la posible confusión (véase capítulo 5).

Tabla 4.7. Principales diferencias entre el sesgo de confusión y la interacción.

SESGO DE CONFUSIÓN	INTERACCIÓN
Es un sesgo	No es un sesgo
Perturbación errónea del efecto	Fenómeno natural
Puede o no estar presente, dependiendo del diseño del estudio	Independiente del diseño del estudio
No se evalúa mediante un test estadístico	Se evalúa mediante un test estadístico
Se puede obtener un estimador global del efecto	No se puede obtener un estimador global del efecto
Se debe evitar	Se debe describir

5. REFERENCIAS BIBLIOGRÁFICAS

COBO E, BUEKENS P. *Necesidad y limitaciones del ajuste.* Med Clin (Barc) 1990;95:702-708.

HULLEY SB, CUMMINGS SM, BROWNER WS, GRADY DG, NEWMAN TB. *Diseño de investigaciones clínicas.* 3ª Ed. Filadelfia: Lippincott, Williams & Wilkins; 2007.

IRALA J, MARTÍNEZ-GONZÁLEZ MA, GUILLÉN GRIMA F. *¿Qué es una variable modificadora del efecto?* Med Clin (Barc) 2001;117(10):297-302.

IRALA J, MARTÍNEZ-GONZÁLEZ MA, GUILLÉN GRIMA F. *¿Qué es una variable de confusión?* Med Clin (Barc) 2001;117(10):377-385.

MANTEL N, HAENSZEL W. *Statistical aspects of the analysis of data from retrospective studies of disease.* J Natl Cancer Inst 1959;22:719-748.

MARTÍNEZ NAVARRO F, ANTÓ JM, CASTELLANOS PL, GILI M, MARSET P, NAVARRO V. *Salud Pública.* Madrid: McGraw-Hill; 1998.

PORTA M. *Dictionary of epidemiology.* 5ª Ed. Nueva York: Oxford University Press; 2008.

ROTHMAN KJ, GREENLAND S. *Modern epidemiolgy.* 2 ed. Filadelfia: Lippincott-Raven; 1998.

SACKETT DL. *Bias in analytic research.* J Chron Dis 1979;32:51-63.

SACKETT DL, HAYNES RB, GUYATT GH, TUGWELL P. *Epidemiología Clínica. Ciencia básica para la medicina clínica.* Madrid: Panamericana; 1995.

SIERRA A, SÁENZ MC, FERNÁNDEZ-CREHUET J, SALLERAS L, CUETO A, et al, editores. *Medicina Preventiva y Salud Pública.* 11ª ed. Barcelona: Elsevier-Masson; 2008.

SZKLO M, NIETO FJ. *Epidemiología Intermedia. Conceptos y Aplicaciones.* Madrid: Ed. Díaz de Santos; 2003.

V. ANÁLISIS DE DATOS EPIDEMIOLÓGICOS

Javier Damián
Nuria Aragonés

1. INTRODUCCIÓN

El análisis consiste básicamente en la obtención de medidas resumen de los datos que ayuden a realizar inferencias. Normalmente el interés reside en la estimación de parámetros (puntual y por intervalos de confianza) y en el contraste de hipótesis.

En este capítulo se presentan algunos procedimientos estadísticos aplicados al análisis de datos de estudios epidemiológicos cuyo objetivo es evaluar la asociación entre un determinante (característica, exposición, hábito...) y la ocurrencia de una enfermedad. En primer lugar se aborda el contraste de hipótesis y la estimación de efectos para el caso del análisis crudo o simple. En segundo lugar se considera el análisis estratificado, en el que se tienen en cuenta variables adicionales a las de exposición y efecto.

Los cálculos dependen del modelo estadístico que se utilice. En general, y debido a su facilidad, se utilizarán modelos basados en la distribución normal. Estos métodos, también denominados *aproximados* o asintóticos, son apropiados en el caso de muestras grandes. Con muestras pequeñas, los cálculos deben realizarse con métodos exactos.

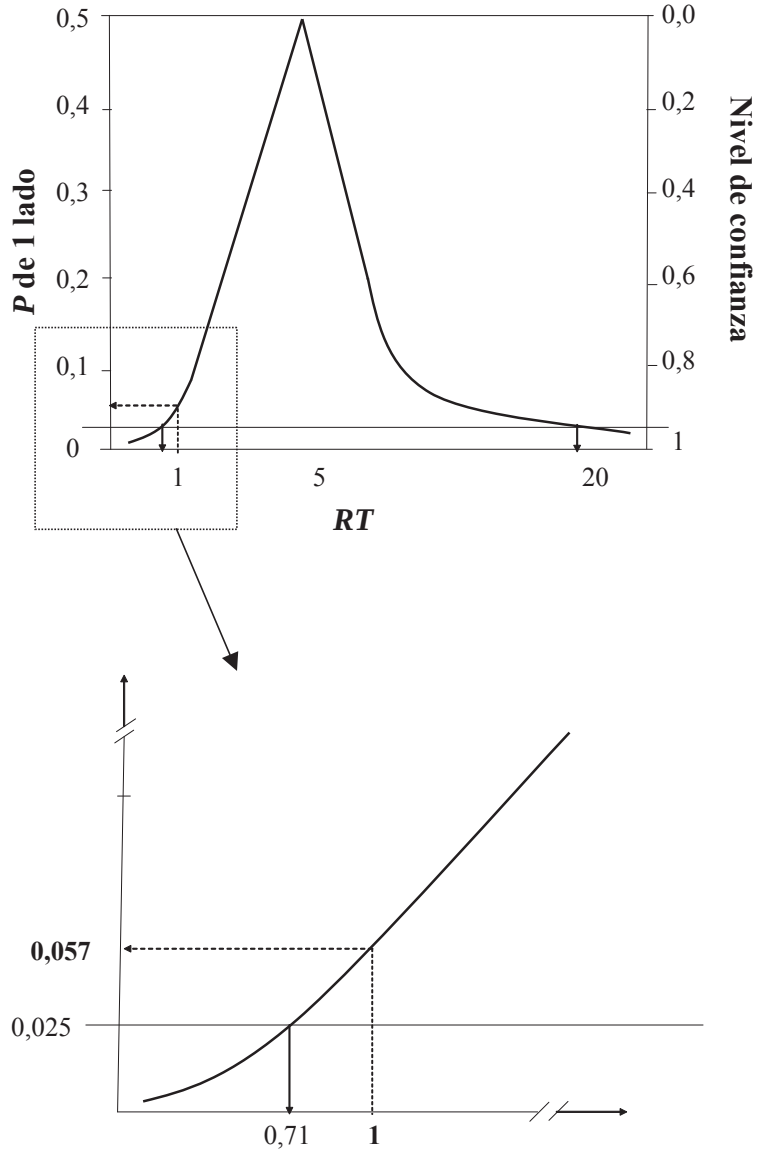
Cuando el objetivo del análisis es obtener información cuantitativa sobre algún parámetro poblacional como una media o un riesgo relativo, se habla de *estimación*, que a su vez puede ser de dos tipos, puntual y por intervalo. Mediante la *estimación puntual* se obtiene un único valor del parámetro desconocido (por ejemplo, un *RR* de 2,43). Una estimación puntual es la mejor estimación numérica de un efecto a partir de nuestros datos, pero dado que es un punto de una escala continua con un número infinito de valores, existen muy pocas posibilidades de que sea enteramente correcta. Por ello necesitamos disponer de una medida de su precisión (del error aleatorio que hay en los datos). La *estimación por intervalo* persigue obtener un rango de los valores más compatibles con el estimador puntual. Los intervalos de confianza se construyen de tal forma que incluyen el verdadero valor con una cierta frecuencia llamada nivel de confianza. El nivel de confianza es seleccionado por el investigador, utilizándose generalmente el 95%.

Por otra parte es habitual que el objetivo del análisis sea *contrastar la hipótesis nula*. Para ello se calcula el valor *P* asociado a la hipótesis nula. Si este valor es menor o mayor de un nivel, preestablecido en el esquema de decisión, se rechaza o no la hipótesis nula.

1.1. Fundamentos del análisis

Los elementos básicos del análisis que se utilizan habitualmente en inferencia se pueden obtener a través de la llamada 'función de valores *P*'. Se trata de una distribución de valores que estiman la compatibilidad del resultado obtenido (p.e. un riesgo relativo) con todo el rango de hipótesis del parámetro (Figura 5-1). Esta función de valores *P* además de ofrecer una idea del error aleatorio de la estimación (de la precisión de la estimación), permitirá calcular algunos valores con interés especial para la inferencia, como el valor *P* asociado a la hipótesis nula o los límites de los diferentes intervalos de confianza. Un intervalo de confianza al 100(1-α)% incluye todos los valores para los cuales su valor *P* bilateral excede el valor α del nivel de confianza, es decir, es un rango de los valores más compatibles con el estimador puntual.

Figura 5-1. Función de valores P correspondiente a los datos del ejemplo 5-1. El pico de la curva corresponde al estimador puntual. Zona ampliada: El valor *P* unilateral (exacto) bajo H_0 es 0,057 (0,114 bilateral). Los límites (exactos) del intervalo de confianza al 95% son 0,71 y 18,5.



2. ANÁLISIS CRUDO

Se entiende como análisis crudo aquel que tiene en cuenta solo las variables principales, normalmente exposición y efecto. Este análisis se aplica cuando no es necesario tener en cuenta ningún factor más aparte de la exposición y de la enfermedad de interés, como ocurre a menudo en estudios experimentales en los que la asignación aleatoria suele prevenir la confusión. Además, el análisis crudo debe ser un primer paso antes de realizar análisis que contemplen otras variables, siendo el caso de los estudios no experimentales, que normalmente requieren análisis estratificado o modelaje multivariante para poder tener en cuenta el papel de otros factores además de las variables principales (factores de confusión y variables de interacción).

Para facilitar su exposición, los métodos de análisis que se presentan a continuación se limitan al caso de una variable de exposición y una de resultado, ambas dicotómicas.

2.1. Datos en forma de personas-tiempo

En un estudio de seguimiento con personas-tiempo como unidades básicas de observación la información utilizada en el análisis se presenta en la siguiente tabla:

	Expuestos	**No expuestos**	
Casos	a	b	M_1
Personas-tiempo	N_1	N_0	T

2.1.1. Contraste de hipótesis

Con este tipo de datos puede asumirse como fijado (sin variabilidad aleatoria) el número total de casos, M_1. El modelo estadístico para el contraste de hipótesis de los datos de persona-tiempo es la distribución binomial, siendo el estadístico el número de casos expuestos (a). Si la hipótesis nula fuera cierta, la media (el número de casos expuestos que cabría esperar, el valor esperado para a será M_1 (N_1 / T). La varianza de esta distribución binomial es: M_1 (N_1 / T) (1-[N_1 / T]), o simplificando: $M_1N_1\ N_0/T^2$. Esta distribución puede aproximarse a una normal (si ambos M_1 (N_1 / T) y M_1(1-[N_1 / T]) son >5) con media y varianza iguales a las de la distribución binomial. Por tanto, se puede calcular el estadístico llamado χ bajo la hipótesis nula como:

$$\chi = \frac{\text{Observado} - \text{Esperado}}{\sqrt{\text{Varianza}}} = \frac{a - \dfrac{N_1 M_1}{T}}{\sqrt{\dfrac{N_1 N_0 M_1}{T^2}}}$$

Y se calculan los valores apropiados en las tablas de la distribución normal estándar $N(0, 1)$.

2.1.2. Estimación puntual y por intervalo de confianza

— *Diferencia de tasas* de incidencia: $DT = I_1 - I_0$

La fórmula para obtener el intervalo de confianza del 95%:

$$DT \pm 1{,}96 \cdot \mathrm{EE}(DT)$$

siendo el error estándar de la diferencia de tasas:

$$\mathrm{EE}(DT) = \sqrt{\frac{a}{N_1^2} + \frac{b}{N_0^2}}$$

— *Razón de tasas*: $RT = \dfrac{I_1}{I_0}$

La razón de tasas de incidencia no tiene una distribución muestral simétrica (Figura 5-2) por lo que es conveniente emplear una transformación logarítmica, que sí es normal. De esta forma se calculan límites de confianza simétricos para el logaritmo de la razón de tasas, invirtiendo luego la transformación mediante la función exponencial:

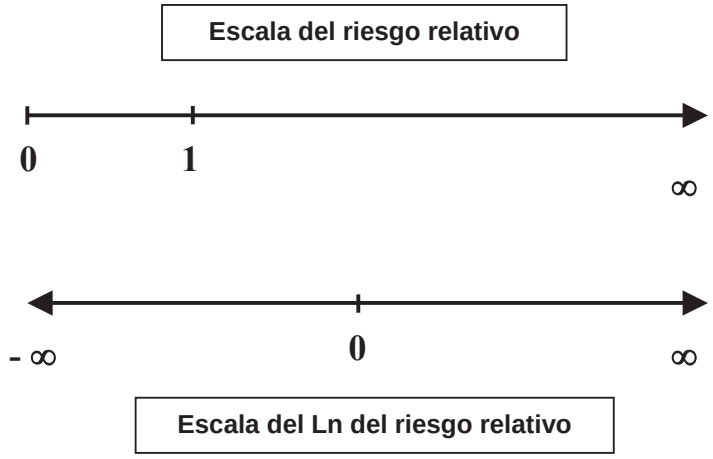
$$\exp[\ln RT \pm 1{,}96 \cdot \mathrm{EE}(\ln RT)] \quad \text{siendo} \quad \mathrm{EE}(\ln RT) = \sqrt{\frac{1}{a} + \frac{1}{b}}$$

Ejemplo 5.1. En el estudio de seguimiento del NHANES II *(National Health and Nutrition Examination Survey*), se han obtenido los siguientes datos de mortalidad cardiovascular en el grupo de edad de 30 a 49 años:

	Hipertensos	Normotensos	**Total**
Casos	3	7	10
Personas-año	2.868	28.309	31.177

DT = **80** por 10^5 personas-año; IC 95%: 80±1,96·61= **-40** a **200**

Figura 5-2. Rango de valores del riesgo relativo en escala original y en escala logarítmica.



$$RT = \frac{3/2.868}{7/28.309} = \mathbf{4{,}23}\ ; \qquad \text{IC 95\%: } \exp[1{,}44 \pm 1{,}96 \cdot 0{,}69] = \mathbf{1{,}09} \text{ a } \mathbf{16{,}36}$$

$$\text{Contraste de hipótesis: } \chi = \frac{\text{Observado} - \text{Esperado}}{\sqrt{\text{Varianza}}} = \frac{3 - \dfrac{2.868 \cdot 10}{31.177}}{\sqrt{\dfrac{2.868 \cdot 28.309 \cdot 10}{31.177^2}}} = \mathbf{2{,}28}$$

que corresponde a un valor *P* bilateral de 0,02. Sin embargo en este ejemplo no se deben utilizar los métodos basados en la aproximación normal ya que M_1 (N_1 / *T*) < 5 (10×2.868/31.177=0,92). El valor *P* calculado por métodos exactos es 0,114 (ver Figura 5-1) con lo que no se rechazaría la hipótesis nula si se utilizara el nivel del 0,05.

2.2. Datos de recuento

La información básica tanto para estudios de cohortes con denominadores en forma de personas en riesgo como para estudios de casos y controles o de prevalencia puede organizarse en forma de una tabla de 2×2, en la que las 4 celdas son frecuencias de sujetos clasificados de

acuerdo con la presencia o ausencia de exposición y de enfermedad, siendo además iguales los métodos de contraste de hipótesis.

	Expuestos	**No expuestos**	
Casos	a	b	M_1
No casos	c	d	M_0
	N_1	N_0	T

2.2.1. Contraste de hipótesis

En este caso igualmente se asumen como fijos los marginales M_1, M_0, N_1, N_0, y el estadístico es igualmente el numero de casos expuestos, a. Bajo hipótesis nula, el valor esperado de a será M_1 (N_1 / T). La varianza es ligeramente diferente a la anterior, ya que el modelo utilizado en este caso es de dos binomiales:

$$\chi = \frac{\text{Observado} - \text{Esperado}}{\sqrt{\text{Varianza}}} = \frac{a - \dfrac{N_1 \cdot M_1}{T}}{\sqrt{\dfrac{N_1 \cdot N_0 \cdot M_1 \cdot M_0}{T^2(T-1)}}}$$

Igualmente se utiliza la distribución normal estándar para obtener los valores P apropiados.

2.2.2. Estimación puntual y por intervalo de confianza

— *Diferencia de riesgos*: $DR = IA_1 - IA_0$

Para conseguir límites aproximados de datos de seguimiento con denominadores compuestos por personas, utilizamos fórmulas ligeramente diferentes a la hora de estimar los errores estándar:

$$DR \pm 1{,}96 \cdot \text{EE}(DR) \qquad \text{EE}(DR) = \sqrt{\frac{a(N_1 - a)}{N_1^3} + \frac{b(N_0 - b)}{N_0^3}}$$

— Razón de riesgos o *riesgo relativo*: $RR = \dfrac{IA_1}{IA_0}$

Como en el caso de las razones de tasas de incidencia, en el caso de la razón de riesgos también es deseable emplear la transformación logarítmica para el cálculo de los intervalos de confianza.

$$\exp[\ln RR \pm 1.96 \cdot \text{EE}(\ln RR)] \qquad \text{EE}(\ln RR) = \sqrt{\frac{c}{a \cdot N_1} + \frac{d}{b \cdot N_0}}$$

— Para datos de casos y controles, la medida de interés central es la odds ratio:

$$OR = \frac{a \cdot d}{b \cdot c}$$

En este caso, el cálculo del error estándar y del intervalo de confianza del 95% se realiza mediante las siguientes operaciones (utilizando también la transformación logarítmica para el cálculo de los intervalos de confianza):

$$\exp[\ln OR \pm 1{,}96 \cdot \mathrm{EE}(\ln OR)] \qquad \mathrm{EE}(\ln OR) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

Tabla 5-1. Análisis crudo de datos. Fórmulas para dos grupos de estudio con muestras grandes (métodos aproximados).

	Datos persona-tiempo	Datos de recuento: personas
Contraste de hipótesis nula	$\chi = \dfrac{a - \dfrac{N_1 \cdot M_1}{T}}{\sqrt{\dfrac{N_1 \cdot N_0 \cdot M_1}{T^2}}}$	$\chi = \dfrac{a - \dfrac{N_1 \cdot M_1}{T}}{\sqrt{\dfrac{N_1 \cdot N_0 \cdot M_1 \cdot M_0}{T^2(T-1)}}}$
Estimación de efectos	• <u>Diferencia de tasas</u> $DT = I_1 - I_0$ $\mathrm{EE}(DT) = \sqrt{\dfrac{a}{N_1^2}+\dfrac{b}{N_0^2}}$ $IC95\%:$ $DT \pm 1.96 \cdot \mathrm{EE}(DT)$ • <u>Razón de tasas (*RT*)</u> $RT = I_1 / I_0$ $\mathrm{EE}(\ln RT) = \sqrt{\dfrac{1}{a}+\dfrac{1}{b}}$ $IC95\%:$ $\exp[\ln RT \pm 1.96 \cdot \mathrm{EE}(\ln RT)]$	• <u>Diferencia de riesgos</u> $DR = IA_1 - IA_0$ $\mathrm{EE}(DR) = \sqrt{\dfrac{a(N_1-a)}{N_1^3}+\dfrac{b(N_0-b)}{N_0^3}}$ $IC95\%:$ $DR \pm 1.96 \cdot \mathrm{EE}(DR)$ • <u>Razón de riesgos (*RR*)</u> $RR = IA_1 / IA_0$ $EE(\ln RR) = \sqrt{\dfrac{c}{a \cdot N_1}+\dfrac{d}{b \cdot N_0}}$ $IC95\%:$ $\exp[\ln RR \pm 1.96 \cdot \mathrm{EE}(\ln RR)]$ • <u>Odds Ratio (*OR*)</u> $OR = a \cdot d / b \cdot c$ $\mathrm{EE}(\ln OR) = \sqrt{\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}}$ $IC95\%:$ $\exp[\ln OR \pm 1.96 \cdot \mathrm{EE}(\ln OR)]$

3. ANÁLISIS ESTRATIFICADO

El análisis estratificado engloba un conjunto de métodos que se basan en la creación de subgrupos (estratos) con el objeto de abordar determinados problemas. Entre las utilidades más habituales se encuentran la identificación y el control de factores de confusión y la evaluación de la modificación del efecto.

La *confusión* es una distorsión en la medida de asociación entre una exposición y una enfermedad resultante del efecto de otra variable que se asocia de forma independiente tanto con la enfermedad como con la exposición.

La *modificación del efecto (o interacción)*, por su parte, se refiere al cambio en la magnitud de la asociación entre la exposición y la enfermedad en función de los valores de una tercera variable (variable modificadora de efecto). También se habla de *heterogeneidad* del efecto.

La estrategia general consiste en crear subgrupos (estratos) formados por las categorías de la tercera variable (confusora o modificadora de efecto) y medir el efecto (riesgo relativo, odds ratio (*OR*), diferencia de tasas...) en esos subgrupos: p.e. calcular el *RR* en hombres y en mujeres; o en los grupos de clase social baja, media y alta. Si se observan diferencias relevantes en el efecto en los estratos es posible que se trate de una variable modificadora de efecto. Si el valor promedio del efecto en los estratos es diferente del valor crudo es posible que se trate de un factor de confusión.

Una de las diferencias más importantes entre confusión y modificación del efecto es que la confusión es un sesgo y, por tanto, hay que identificarlo y controlarlo, por lo que procederá calcular el promedio del efecto uniforme observado en los estratos. La modificación del efecto es un atributo del fenómeno que se estudia, por lo que el objetivo es describirlo y evaluarlo. Si se aprecia modificación de efecto lo esencial es describir el patrón observado en los estratos, teniendo poco sentido calcular un estimador promedio.

La obtención de estimadores ajustados o "conjuntados" ("*pooled*") se basa en el cálculo del promedio ponderado de los estimadores de los estratos. Existen varios métodos en función de la elección de las ponderaciones que se asignarán a cada estrato. Aquí se presentan los tres métodos más habituales: la estandarización, el método de *Woolf*, y el método de *Mantel y Haenszel*.

La *estandarización* utiliza unos pesos comunes para los grupos que se comparan que pueden ser incluso externos a los datos. El *método de Woolf* asigna un peso proporcional al inverso de la varianza en cada estrato: estratos con gran error aleatorio (menos informativos) aportarán menos al estimador conjuntado ('pesarán menos'). El *método de Mantel-Haenszel* utiliza un grupo *especial* de pesos basados en los datos de cada estrato, con un componente igualmente proporcional al error aleatorio. A continuación se describen estos métodos.

3.1. Estandarización

La estandarización es un procedimiento de control de factores de confusión que se basa en la sustitución de las distribuciones de los factores de confusión de las poblaciones que se comparan por otras distribuciones estándar (o comunes).

Grupo de edad	Nº de casos N_i	Personas-tiempo P_i	Tasa de Incidencia ($I_i = N_i / P_i$)
15-49	N_1	P_1	I_1
50-64	N_2	P_2	I_2
65 +	N_3	P_3	I_3
Total	N_t	P_t	I_t

La tasa de incidencia total (o "tasa cruda") es una *media ponderada* de las tasas específicas por subgrupos. En el caso de la edad:

$$I_t = \frac{N_t}{P_t} = \frac{\sum I_i P_i}{\sum P_i}$$

Si se sustituyen los pesos P_i (la distribución por edad) de esta población por los de otra población (w_i) – normalmente llamada "estándar" o de referencia— se obtiene una *tasa estandarizada:*

$$I_E = \frac{\sum w_i I_i}{\sum w_i}$$

Una tasa estandarizada es una medida hipotética. Corresponde a la "tasa cruda" que tendría la población de estudio si en lugar de tener su distribución por edad tuviera la distribución por edad estándar. Por tanto esta tasa hipotética no tiene utilidad si no es para ser comparada con otras tasas que hayan sido estandarizadas utilizando la misma distribución por edad.

Si calculamos la tasa estandarizada para *otra población* utilizando el *mismo grupo de pesos* (la misma distribución estándar) las diferencias entre las dos tasas estandarizadas no podrán ser atribuidas a una diferente distribución por edad. Así, las tasas estandarizadas de dos poblaciones, una de expuestos (I_1) y otra de no expuestos (I_0), utilizando la misma distribución por edad (w_i), serán:

$$I_{1E} = \frac{\sum w_i I_{1i}}{\sum w_i} \qquad I_{0E} = \frac{\sum w_i I_{0i}}{\sum w_i}$$

Nótese que ambas tasas estandarizadas han utilizado el mismo grupo de pesos (w_i). Estos pesos pueden venir de diversas fuentes: de la distribución de una de las dos poblaciones que se comparan, de la suma de las dos poblaciones, o de otra distribución (real o no). En principio lo importante es que la distribución sea la misma para ambas poblaciones.

Para obtener medidas de efecto estandarizadas por edad, se calcula la razón o la diferencia entre las dos tasas estandarizadas:

$$RT_E = \frac{\sum w_i I_{1i}}{\sum w_i} \Big/ \frac{\sum w_i I_{0i}}{\sum w_i} = \frac{\sum w_i I_{1i}}{\sum w_i I_{0i}} = \frac{\sum w_i I_{0i} RT_i}{\sum w_i I_{0i}}$$

$$DT_E = \frac{\sum w_i I_{1i}}{\sum w_i} - \frac{\sum w_i I_{0i}}{\sum w_i} = \frac{\sum w_i DT_i}{\sum w_i}$$

En estos ejemplos se ha utilizado la edad como variable de estandarización pero el mismo proceso puede utilizarse para estandarizar por cualquier otra variable: por sexo, por clase social; por edad y sexo; por edad, sexo y clase social...

Ejemplo 5-2: Mortalidad por enfermedad cardiovascular en el estudio de seguimiento NHANES II. 7.801 personas libres de enfermedad cardiovascular basal seguidas un promedio de 15 años.

	Hipertensos			Normotensos (<140 mmHg)	
Edad	casos	Personas-año		casos	Personas-año
30-49	3	2.868		7	28.309
50-69	80	13.867		67	31.016
70+	253	14.842		172	15.348
Total	336	31.577		246	74.673

Las tasas crudas son:

I_1= 336/31.577 = 1.064 /10^5/año $\qquad$ I_0= 246/74.673 = 329 /10^5/año

Y la razón de tasas crudas: $\boldsymbol{RT = 3{,}2}$

Las tasas específicas por edad junto con las razones de tasas específicas son:

$I_{1\ 30\text{-}49}$ = 3/2.868 = 105 $\qquad$ $I_{0\ 30\text{-}49}$ = 7/28.309 = 25 ($RT_{30\text{-}49}$ = 4,2)
$I_{1\ 50\text{-}69}$ = 80/13.867 = 577 $\qquad$ $I_{0\ 50\text{-}69}$ = 67/31.016 = 216 ($RT_{50\text{-}69}$ = 2,7)
$I_{1\ 70+}$ = 253/14.842= 1.705 $\qquad$ $I_{0\ 70+}$ = 172/15.348 = 1.121 (RT_{70+} = 1,5)

Tomando como grupo de pesos estándar la suma de ambas poblaciones:

$$w_{\ 30\text{-}49} = 2.868 + 28.309 = 31.177$$
$$w_{\ 50\text{-}69} = 13.867 + 31.016 = 44.883$$
$$w_{\ 70+} = 14.842 + 15.348 = 30.190$$

Las tasas estandarizadas son:

$$I_{1E} = \frac{\sum w_i I_{1i}}{\sum w_i} = \frac{(105 \cdot 31.177) + (577 \cdot 44.883) + (1.705 \cdot 30.190)}{31.177 + 44.883 + 30.190} = 759\ /10^5/\text{año}$$

$$I_{0E} = \frac{\sum w_i I_{0i}}{\sum w_i} = \frac{(25 \cdot 31.177) + (216 \cdot 44.883) + (1.121 \cdot 30.190)}{31.177 + 44.883 + 30.190} = 417\ /10^5/\text{año}$$

y la razón de tasas estandarizadas: $\boldsymbol{RT_E = 1{,}8}$

Que se interpreta como la razón de tasas esperable si ambas poblaciones tuvieran la misma distribución por edad (la distribución estándar). En todo caso hay que recordar la importancia de describir cualquier patrón observado en los estratos. En este ejemplo se observa con claridad que el efecto de la exposición es importante en los más jóvenes y muestra una tendencia decreciente con la edad.

3.1.1. Razón de mortalidad estandarizada

Cuando se utiliza como estándar la distribución de una de las dos poblaciones que se comparan (normalmente la de la población expuesta) la tasa cruda de la población expuesta

coincide con la tasa estandarizada por lo que la razón de tasas estandarizadas se convierte en la razón entre un número de *casos observados* (en la población expuesta) y un número de *casos esperados* (los casos que sería esperable obtener en la población no expuesta si su distribución por edad fuera la de la población expuesta). Esta razón ha sido denominada *Razón de Mortalidad (o morbilidad) Estandarizada (RME):*

$$RME = \frac{\sum w_i I_{1i}}{\sum w_i} \Big/ \frac{\sum w_i I_{0i}}{\sum w_i} = \frac{\sum w_i I_{1i}}{\sum w_i I_{0i}} = \frac{\text{casos } observados}{\text{casos } esperados}$$

Normalmente dos o más *RME*s no son comparables entre sí debido a que las distribuciones estándar son diferentes en el cálculo de cada *RME,* aunque en la mayoría de la situaciones el sesgo que se produce es pequeño. Por otra parte el error aleatorio es menor por lo que habrá situaciones en las que será preferible utilizar *RME*s.

La estrategia consistente en utilizar unas 'tasas estándar' para obtener la *RME* suele denominarse *estandarización 'indirecta'*, aunque el procedimiento es idéntico al anterior (también llamado *estandarización 'directa'*). Por ejemplo, la mortalidad por cáncer en trabajadores se compara con la correspondiente a la población general activa, que aporta las 'tasas estándar', calculándose la razón de tasas estandarizadas utilizando la distribución por edad de los trabajadores como estándar.

Ejemplo 5-3: En un estudio se analiza la incidencia de cáncer de mama en diversas ocupaciones (Pollán y Gustavsson) y se comparan con la población activa de mujeres. En el caso de las médicas se han obtenido los siguientes datos:

	Médicas			Población de referencia
Edad	casos	Personas-año		Tasas año^{-1} ($I_R i$)
25-44		11.554		0,00052
45-59		12.878		0,00166
60-79		4.994		0,00238
Total	59			

$w_{25\text{-}44} = 11.554$; $w_{45\text{-}59} = 12.878$; $w_{60\text{-}79} = 4.994$

$$RME = \frac{\sum w_i I_{1i}}{\sum w_i} \Big/ \frac{\sum w_i I_{Ri}}{\sum w_i} = \frac{\sum w_i I_{1i}}{\sum w_i I_{Ri}} = \frac{\text{casos } observados}{\text{casos } esperados} =$$

$$= \frac{59}{(0{,}00052 \cdot 11.554) + (0{,}00166 \cdot 12.878) + (0{,}00238 \cdot 4.994)} = \frac{59}{39} = \mathbf{1{,}50}$$

La incidencia en el grupo de médicas es un 50% mayor que en la población activa de mujeres, si ambos grupos tuvieran la misma distribución por edad (la de las médicas). Como se aprecia, el método empleado es el mismo que anteriormente excepto por el grupo de pesos estándar.

En el caso de las enfermeras y auxiliares la *RME* fue de 0,99. En teoría no se puede realizar una comparación directa con las médicas ya que ambas *RME*s han sido obtenidas con distribuciones estándar diferentes (quizás no muy diferentes).

3.2. Método de *Woolf*

Supongamos que queremos evaluar el efecto de una variable de exposición principal dicotómica (expuestos y no expuestos) sobre el desarrollo de una enfermedad (casos de la enfermedad y controles) utilizando la odds ratio (*OR*) como medida de asociación. Se desea valorar el papel de una tercera variable que es un potencial factor de confusión o de modificación de efecto.

Ejemplo 5-4. En el estudio EURAMIC, queremos evaluar el efecto de tener un índice cintura/cadera, WHR (Waist to hip ratio) por encima de la mediana (expuestos: WHR por encima de la mediana; no expuestos: WHR por debajo de la mediana) sobre el riesgo de infarto de miocardio (casos y controles), controlando el efecto del consumo de tabaco, agrupado en 3 categorías: nunca fumadores; ex-fumadores y fumadores actuales.

La variable de estratificación es siempre categórica; si se quiere estratificar por una variable continua, es necesario agruparla en categorías. Asimismo, si queremos estratificar por más de una variable, es necesario crear tantos estratos como combinaciones de las categorías de las variables de estratificación: si el primer factor de estratificación tiene, por ejemplo, 3 categorías y el segundo 4, será necesario crear 12 estratos. El análisis estratificado puede sistematizarse en las siguientes etapas:

1. Dividir la muestra en estratos exhaustivos y mutuamente excluyentes.
2. Calcular la medida de asociación específica para cada estrato (en este ejemplo utilizaremos la *OR*, aunque el proceso es análogo para el cálculo de riesgos relativos y de diferencias de riesgos).
3. Comprobar si las medidas de efecto son similares en todos los estratos mediante la realización de un *test de homogeneidad*. Si los resultados en los estratos son diferentes (es decir, si la variable de estratificación *modifica el efecto* de la variable de exposición principal), se describirán las medidas de asociación en cada estrato como resultado final del análisis. Si los resultados en los estratos son lo suficientemente homogéneos se entenderá que en todos los estratos se está estimando una *OR* única y que las diferencias encontradas son debidas a variabilidad aleatoria. En este caso procederemos a combinarlas, obteniendo un estimador ajustado de la *OR*, libre de confusión por esa variable ya que se calcula a partir de los estimadores específicos de cada estrato, que a su vez están libres de confusión. Posteriormente se calcula el intervalo de confianza para la *OR* (y, si se desea, el test de significación).

Notación:

	Expuestos	No expuestos	Total
Casos	a_k	b_k	M_{1k}
Controles	c_k	d_k	M_{0k}
Total	N_{1k}	N_{0k}	T_k

A partir de esta tabla se calcula la *OR* del estrato *k* (OR_k), su logaritmo natural ($\ln OR_k$) y la varianza del logaritmo natural ($\mathrm{Var}(\ln OR_k)$). Debido a la distribución asimétrica de la *OR* es preferible realizar los cálculos en la escala logarítmica:

— $OR_k = \dfrac{a_k \cdot d_k}{b_k \cdot c_k}$

— $\ln OR_k$

— $\mathrm{Var}(\ln OR_k) = \dfrac{1}{a_k} + \dfrac{1}{b_k} + \dfrac{1}{c_k} + \dfrac{1}{d_k}$

La varianza del estrato *k*, está relacionada con la 'información' que aporta ese estrato: cuanto mayor sea var(lnOR_k) menor será la información que nos aporta (es decir, más incertidumbre tiene nuestra estimación). Este concepto se formaliza definiendo el 'peso' del estrato *k* (w_k), como el inverso de la varianza para ese estrato:

$$w_k = \frac{1}{\mathrm{Var}(\ln OR_k)}$$

Ejemplo 5-4: En la evaluación de la asociación entre el WHR y el riesgo de infarto de miocardio en el estudio EURAMIC, se obtuvo la siguiente tabla de 2×2 cruda:

	WHR alto	WHR bajo
Casos	386	124
Controles	301	299

$$OR_c = \frac{386 \cdot 299}{124 \cdot 301} = 3{,}09$$

Para cada uno de los 3 estratos de hábito tabáquico se obtienen las siguientes tablas de 2×2:

Estrato 1: No fumadores

	WHR alto	WHR bajo
Casos	39	17
Controles	65	86

— $OR_1 = \frac{39 \cdot 86}{17 \cdot 65} = 3{,}04$

— $\ln OR_1 = \ln 3{,}04 = 1{,}11$

— $\mathrm{Var}(\ln OR_1) = \frac{1}{39} + \frac{1}{65} + \frac{1}{17} + \frac{1}{86} = 0{,}111$

— $w_1 = \frac{1}{0{,}111} = 8{,}97$

Estrato 2: Ex fumadores

	WHR alto	WHR bajo
Casos	108	27
Controles	110	102

— $OR_2 = \frac{108 \cdot 102}{27 \cdot 110} = 3{,}71$

— $\ln OR_2 = \ln 3{,}71 = 1{,}31$

— $\mathrm{Var}(\ln OR_2) = \frac{1}{108} + \frac{1}{110} + \frac{1}{27} + \frac{1}{102} = 0{,}065$

— $w_2 = \frac{1}{0{,}065} = 15{,}34$

Estrato 3: Fumadores

	WHR alto	WHR bajo
Casos	239	80
Controles	126	111

— $OR_3 = \dfrac{239 \cdot 111}{80 \cdot 126} = 2{,}63$

— $\ln OR_3 = \ln 2{,}63 = 0{,}97$

— $\mathrm{Var}(\ln OR_3) = \dfrac{1}{239} + \dfrac{1}{126} + \dfrac{1}{80} + \dfrac{1}{111} = 0{,}034$

— $w_3 = \dfrac{1}{0{,}034} = 29{,}736$

Tabla resumen:

	OR_k	$\ln OR_k$	Var($\ln OR_k$)	w_k [1/ Var($\ln OR_k$)]
No fumadores	3,04	1,11	0,111	9,0
Ex-fumadores	3,71	1,31	0,065	15,3
Fumadores	2,63	0,97	0,034	29,7

3.2.1. Análisis de homogeneidad de los estimadores de efecto de los distintos estratos

La evaluación de la homogeneidad inter-estrato consiste en determinar hasta qué punto las diferencias observadas en las *OR*s son compatibles con la hipótesis (nula) de que son iguales. Para contrastar esta hipótesis es necesario realizar un *test de homogeneidad* entre los estratos. El test de homogeneidad de *Woolf* evalúa la diferencia entre las *OR*s específicas de cada estrato y su media: cuanto más heterogéneas sean las *OR*s, más alejadas estarán de su media y mayor será el estadístico de homogeneidad. Los pasos detallados para calcular este estadístico se presentan a continuación:

Calcular la media ponderada de los logaritmos naturales de las *OR*s de cada estrato, utilizando como factores de ponderación los pesos de cada estrato:

$$\overline{\ln OR}^{w} = \frac{\sum_k w_k \cdot \ln OR_k}{\sum_k w_k}$$

Calcular el *estadístico de homogeneidad*:

$$\chi^2_{k-1} = \sum_k \frac{(\ln OR_k - \overline{\ln OR}^{w})^2}{\mathrm{Var}(\ln OR_k)}$$

Para evaluvcar la hipótesis nula de homogeneidad de las *OR*s específicas (H_0: $OR_1 = OR_2 = \ldots = OR_k$), contrastar el valor obtenido con los percentiles de la distribución χ^2 con k-1 grados de libertad, obtener el valor P y decidir en función del nivel de significación elegido.

Ejemplo. La evaluación de la homogeneidad entre las distintas categorías de hábito tabáquico en la asociación entre WHR e infarto agudo de miocardio daría como resultado:

$$\overline{\ln OR}^{w} = \frac{8{,}97 \cdot 1{,}11 + 15{,}34 \cdot 1{,}31 + 29{,}74 \cdot 0{,}97}{8{,}97 + 15{,}34 + 29{,}74} = 1{,}09$$

$$\chi^2 = \frac{(1{,}11 - 1{,}09)^2}{0{,}111} + \frac{(1{,}31 - 1{,}09)^2}{0{,}065} + \frac{(0{,}97 - 1{,}09)^2}{0{,}034} = 1{,}2$$

Este valor se ha de contrastar con los percentiles de la distribución χ^2 con k - 1 = 2 grados de libertad, lo que da un valor P = 0,55. Este valor es bastante compatible con la hipótesis nula de homogeneidad. Si se establece un nivel de significación de, por ejemplo 0,15, el resultado es *no* significativo: No admitimos interacción entre tabaco y el índice cintura/cadera en el riesgo de desarrollar infarto de miocardio, asumiendo por tanto que las diferencias entre los estratos se deben a variabilidad aleatoria.

Si se encuentra interacción, el análisis puede concluirse presentando los estimadores de cada estrato y sus intervalos de confianza, junto con el valor P del test de homogeneidad. Hay que resaltar que los test estadísticos de homogeneidad son poco potentes, por lo que si se aprecia algún patrón evidente puede ser conveniente describirlo (por ejemplo algún orden o tendencia).

3.2.2. Obtención de un estimador ajustado de la OR

Si no existe interacción, el proceso final del análisis estratificado consiste en la obtención de un estimador global, no influenciado por el efecto del factor de confusión (es decir, ajustado por el factor de confusión).

Anteriormente se obtuvo $\overline{\ln OR}^{w}$ como la media ponderada de los logaritmos naturales de las *OR*s específicas. El estimador de la *OR* ajustada por el método de *Woolf* se calcula como:

$$OR_a = \exp(\overline{\ln OR}^{w})$$

El error estándar de $\overline{\ln OR}^{w}$:

$$\mathrm{EE}(\overline{\ln OR}^{w}) = \sqrt{\frac{1}{\sum_k w_k}}$$

Estas fórmulas permiten calcular un intervalo de confianza para OR_a, así como un valor P para contrastar la H_0 de que $OR_a = 1$.

— *IC* del 95% para OR_a:

$$\exp[\overline{\ln OR}^{w} \pm 1{,}96 \cdot \mathrm{EE}(\overline{\ln OR}^{w})]$$

Para el contraste de hipótesis para H_0: $OR_a = 1$, calcular

$$z = \frac{\overline{\ln OR}^{w}}{\mathrm{EE}(\overline{\ln OR}^{w})}$$

y contrastar el resultado obtenido con los percentiles de la distribución normal estandarizada, N (0, 1).

Ejemplo: Siguiendo con el ejemplo de estudio EURAMIC, la *OR* cruda y la ajustada difieren poco (OR_c = 3,09 y OR_a = exp(1,09) = 2,97). El consumo de tabaco no parece ser un factor de confusión importante para la asociación entre el índice cintura/cadera y el riesgo de infarto de miocardio en este estudio. El intervalo de confianza al 95% para la *OR* ajustada es: exp[1,09 ± (1,96·0,136)] = 2,28-3,88, y el estadístico z = 1,09/0,136 = 8,01, que da un valor P < 0,0001.

3.3. Método de Mantel-Haenszel

Mantel y Haenszel propusieron una fórmula sencilla de calcular la *OR* ajustada. Se considera el mejor método en la mayoría de las situaciones:

$$OR_{MH} = \frac{\sum_k \frac{a_k \cdot d_k}{T_k}}{\sum_k \frac{b_k \cdot c_k}{T_k}}$$

El error estándar del $\ln OR_{MH}$ (Robins et al):

$$\mathrm{EE}(\ln OR_{MH}) = \sqrt{\frac{\sum_k P_k R_k}{2\left(\sum_k R_k\right)^2} + \frac{\sum_k (P_k S_k + Q_k R_k)}{2\left(\sum_k R_k\right)\left(\sum_k S_k\right)} + \frac{\sum_k Q_k S_k}{2\left(\sum_k S_k\right)^2}}$$

donde $P_k = \frac{a_k + d_k}{T_k}$, $Q_k = \frac{b_k + c_k}{T_k}$, $R_k = \frac{a_k \cdot d_k}{T_k}$ y $S_k = \frac{b_k \cdot c_k}{T_k}$.

Así, el *intervalo de confianza* para OR_{MH} y la prueba de hipótesis para H_0: $OR_{MH} = 1$ se calculan como:

— *IC* 95% OR_{MH}: $\exp\left[\ln OR_{MH} \pm 1{,}96 \cdot \mathrm{EE}(\ln OR_{MH})\right]$

— Para el contraste de la hipótesis nula H_0: $OR_{MH} = 1$, se calcula:

$$z = \frac{\ln OR_{MH}}{\mathrm{EE}(\ln OR_{MH})}$$

Ejemplo: La estimación de la *OR* combinada por el método de *Mantel-Haenszel* en el ejemplo es:

$$OR_{MH} = \frac{\frac{39 \cdot 86}{207} + \frac{108 \cdot 102}{347} + \frac{239 \cdot 111}{556}}{\frac{17 \cdot 65}{207} + \frac{27 \cdot 110}{347} + \frac{80 \cdot 126}{556}} = 2{,}99$$

que es similar a la obtenida por el método de *Woolf* (OR_a = 2,97).

El error estándar del lnOR_{MH}: 0,136, con el que podemos calcular intervalos de confianza y valores *P*: *IC* 95% OR_{MH} = 2,29 – 3,90; z = 8,09, *P* < 0,0001.

3.4. Fórmulas de Mantel-Haenszel y Miettinen

Para contrastar la hipótesis nula (H_0: $OR_{MH} = 1$) se puede utilizar la fórmula de χ *de Mantel-Haenszel*, que bajo la hipótesis nula sigue una distribución normal estandarizada. Esta fórmula es útil además para calcular posteriormente los intervalos de confianza de *Miettinen (basados en el test).*

$$\chi_{MH} = \frac{\sum_k a_k - \sum_k \frac{N_{1k} \cdot M_{1k}}{T_k}}{\sqrt{\sum_k \frac{N_{1k} \cdot M_{1k} \cdot N_{0k} \cdot M_{0k}}{T_k^2 \cdot (T_k - 1)}}}$$

Para calcular los intervalos de confianza de *Miettinen*:

$$OR_{MH}^{\left(1 \pm \frac{z_{\alpha/2}}{\chi_{MH}}\right)}$$

En el *ejemplo*: χ_{MH} = 8,19; *P* < 0,0001. Y el *IC*95% de *Miettinen* = 2,30 - 3,88

Los intervalos de confianza de *Miettinen* son aproximados pero son bastante exactos para valores no muy alejados del nulo.

3.5. Estimadores de Mantel-Haenszel en otros diseños

Existen formulas equivalentes para otros estimadores combinados de *Mantel-Haenszel* (diferencia de tasas, razón de tasas, diferencia de riesgos, razón de riesgos) y sus varianzas. Se recomienda consultar estas fórmulas en Greenland y Rothman.

3.6. Diferencias entre los métodos de análisis estratificado.

Los métodos de cálculo de estimadores ajustados de *Woolf* y de *Mantel-Haenszel* asumen homogeneidad del efecto. Esta asunción no es tan clara en el caso de la estandarización ya que la elección de los pesos puede ser arbitraria. Si se aprecia modificación del efecto lo esencial es presentar y describir los estimadores de los estratos, pero si además se desea obtener una medida global del efecto de neto de la exposición es preferible obtener estimadores combinados utilizando la estandarización (también en los diseños de casos y controles se puede utilizar estandarización. Ver Greenland y Rothman).

En cuanto a las diferencias entre el método de *Woolf* y el de *Mantel-Haenszel* cabe resaltar que el método de *Woolf* requiere que todas las celdas de cada estrato incluyan alto número de efectivos; si no es así, como ocurre en la mayoría de las ocasiones, resulta fácil que las estimaciones sea erróneas. Esta limitación es el principal problema del análisis estratificado: cuando el número de factores de confusión es alto los estratos van teniendo menos efectivos y las celdas van quedando vacías. Supongamos, en una situación muy común, que disponemos de un tamaño muestral de 800 personas en un estudio de casos y controles, y que deseamos controlar por edad en 3 categorías, por sexo en 2, por clase social en 3, y por consumo de tabaco en 3. El número de estratos es de: 3×2×3×3 = 54. En el caso (por otra parte muy raro) de que las 800 personas se distribuyeran equitativamente en los estratos, éstos tendrían de promedio 14,8 personas. Lo más habitual es que hubiera algunos estratos sin casos o sin controles. Si además consideramos conveniente ajustar por hipertensión en dos categorías, el número total será de: 54×2 = 108, reduciéndose el promedio a 7,4. Este problema es menor si se utiliza el método de

Mantel-Haenszel, que permite obtener estimaciones válidas con muy pocos efectivos en cada estrato. Esta y otras propiedades hacen de éste el *método de elección*, añadiendo a sus ventajas la facilidad de cálculo. Aún así, en muchas situaciones el número de estratos en relación con el tamaño de la muestra será excesivo y se recurrirá entonces a la utilización de modelos de regresión (*regresión logística* en el caso de variables respuesta categóricas). Aún en el caso de que sea finalmente necesario utilizar modelos de regresión, es muy conveniente realizar análisis estratificado incluyendo aquellas variables con mayor importancia, bien científica o de salud pública, ya que esto permite "conocer" mejor los datos, lo cual es más complicado en el caso de modelos de regresión multivariantes.

3.7. Confusión residual

Se habla de confusión residual cuando por alguna razón no se ha controlado toda la potencial confusión en un estudio, bien por que no se ha ajustado por alguna/as variables de confusión o porque se categorizan variables de confusión continuas en pocos estratos por lo que no se puede asumir ausencia de confusión dentro de cada estrato. Por ejemplo, si se desea controlar el efecto de la edad en una población entre 15 y 65 años y se crean sólo dos estratos (p.e.: 15-35 años y 36-65) puede producirse confusión dentro de cada estrato (el caso extremo es *un* solo estrato, es decir datos crudos). Normalmente con pocos estratos se controla bien la confusión en la mayoría de las situaciones, siendo la estrategia más adecuada comprobar cual es el número mínimo de estratos que garantice un nivel de control que se considere adecuado.

3.8. Aplicación al análisis de datos emparejados

En los estudios de casos y controles emparejados normalmente es necesario controlar por las variables de emparejamiento ya que si no es así se introducirá un sesgo. El análisis de un estudio de casos y controles emparejado se realiza estratificando por parejas; es decir cada pareja caso-control forma un estrato. El análisis entonces se realiza mediante la fórmula de *Mantel-Haenszel*, apropiada para datos esparcidos. Los estratos pueden clasificarse en 4 tipos: dos en los que la pareja es concordante (T y W) en relación con la exposición y dos tipos en los que son discordantes (U y V):

	T		*U*		*V*		*W*	
	caso	control	caso	control	caso	control	caso	control
Exp	1	1	1	0	0	1	0	0
No exp	0	0	0	1	1	0	1	1

$$\frac{a_i \cdot d_i}{T_i} = \frac{b_i \cdot c_i}{T_i} = 0 \qquad \frac{a_i \cdot d_i}{T_i} = 1/2 \qquad \frac{b_i \cdot c_i}{T_i} = 1/2 \qquad \frac{a_i \cdot d_i}{T_i} = \frac{b_i \cdot c_i}{T_i} = 0$$

Los estratos concordantes no aportan información a la *OR* combinada por lo que la fórmula se simplifica a la razón entre el número de parejas discordantes en la que el caso está expuesto (*U*) y en la que lo está el control (*V*):

$$OR_{MH} = \frac{\sum \frac{a_i \cdot d_i}{T_i}}{\sum \frac{b_i \cdot c_i}{T_i}} \qquad OR_{MH} = \frac{U \cdot 1/2}{V \cdot 1/2} = \frac{U}{V}$$

4. REFERENCIAS BIBLIOGRÁFICAS

GREENLAND S, ROTHMAN KJ. *Introduction to stratified analysis.* En: Rothman KJ, Greenland S. Modern Epidemiology. Filadelfia: Lippincott-Raven, 1998.

HENNEKENS CH, BURING JE. *Analysis of epidemiologic studies: Evaluating the role of confounding.* En: Epidemiology in Medicine. Boston: Little, Brown and Company, 1987.

KAHN HA, SEMPOS CT. *Statistical Methods in Epidemiology.* Nueva York: Oxford University Press, 1989

KLEINBAUM DG, KUPPER LL, MORGENSTERN H. *Stratified analysis.* En: Epidemiologic Research. Principles and Quantitative Methods. Nueva York: Van Nostrand Reinhold, 1982.

MANTEL N, HAENSZEL WH. *Statistical aspects of the analysis of data fron retrospective studies of disease*. *J Natl Cancer Inst* 1959; 22: 719-748.

MARTÍN ANDRÉS A, LUNA DEL CASTILLO JD. *Bioestadística para las Ciencias de la Salud.* Madrid: Ediciones Norma S. A., 1994.

POLLÁN M, GUSTAVSSON P. *High-risk occupations for breast cancer in the Swedish female working population. Am J Public Health* 1999; 89: 875-881.

ROBINS J, GREENLAND S, BRESLOW N. *A general estimation for the variance of the Mantel-Haeszel odds ratio. Am J Epidemiol* 1986; 124: 719-723.

SZKLO M, NIETO FJ. *Stratification and adjustment: Multivariate analysis in epidemiology.* En: Epidemiology beyond the basics. Gaithersburg: Aspen, 2000.

VI. ESTUDIOS DE COHORTES

Marina Pollán
Beatriz Pérez-Gómez

Los estudios de cohortes se caracterizan por la identificación y seguimiento de uno o más grupos de individuos con un determinado nivel de exposición, para detectar y cuantificar la aparición del evento o enfermedad de interés a lo largo del tiempo.

La palabra "cohorte" deriva del vocablo latino *cohors – cohortis,* que designaba a una unidad táctica de infantería dentro de las legiones romanas. En epidemiología, el término "cohorte" fue incorporado en 1935 y sirve en la actualidad para referirse a un conjunto de personas que comparten una determinada exposición y/o "evolucionan" juntas a lo largo del tiempo.

La esencia de los estudios de cohortes es el seguimiento. En el diseño de cohortes está implícita la recogida de información, a nivel individual, sobre la exposición de uno o varios grupos de personas (cohortes), y el establecimiento de un mecanismo para seguir a los miembros de estas cohortes con objeto de detectar la aparición del evento de interés y también la posible salida o abandono del estudio. El seguimiento sirve para cuantificar no sólo la frecuencia de enfermedad en las distintas cohortes (incidencia acumulada), sino el ritmo o dinámica de la enfermedad en cada una de ellas (tasa de incidencia). Además, durante el seguimiento es posible considerar e incorporar en el análisis los cambios en el nivel de exposición así como las variaciones del resto de factores de riesgo relevantes.

Por tanto, los estudios de cohortes valoran el estado de exposición antes de que se produzca la enfermedad y permiten registrar los cambios ocurridos en el tiempo a nivel individual. La mayor parte de las enfermedades crónicas son el resultado de un proceso que se extiende a lo largo de décadas, en el que intervienen diferentes exposiciones o factores de riesgo. La naturaleza dinámica de la exposición a estos factores y su interrelación a lo largo del tiempo sólo puede ser estudiada adecuadamente en un diseño de cohortes. Esta dinámica temporal está ausente de los estudios transversales y es difícilmente investigable en los estudios de casos y controles.

1. ALGUNOS EJEMPLOS DE ESTUDIOS DE COHORTES

A continuación se presentan 6 estudios de cohortes que servirán de ejemplo para ilustrar los aspectos básicos del diseño desarrollados más adelante.

Ejemplo 1: Cohorte de médicos británicos

El estudio prospectivo de los efectos del tabaco sobre la salud comenzó en 1951. En 1964 los autores publicaron un artículo con el diseño de su estudio y los resultados correspondientes a los primeros diez años de seguimiento. El estudio invitaba a participar a los médicos de Reino Unido, rellenando un cuestionario que fue enviado el 31 de octubre de 1951 a 59.600 médicos de ambos sexos que figuraban en el Registro de Médicos. Los investigadores optaron por un cuestionario corto que favoreciese la participación. Se pidió a los médicos que se clasificasen como: fumadores actuales, exfumadores o no fumadores (definidos como aquellos que no han llegado a fumar un cigarrillo al día durante 1 año). A los fumadores o exfumadores se les pidió que informasen sobre la edad de inicio y la cantidad de cigarrillos/día que fumaban en ese momento o antes de abandonar el hábito.

Un total de 40.637 médicos (34.445 hombres y 6.192 mujeres) contestaron el cuestionario inicial. Durante el seguimiento se les enviaron otros tres cuestionarios a los hombres y otros dos a las mujeres para estimar los cambios en el patrón de consumo.

El seguimiento se realizó mediante el Registro Nacional de Defunción. En 1964 se publicaron los resultados de los diez primeros años de seguimiento y posteriormente los correspondientes a veinte y veintidós años de seguimiento en hombres y en mujeres respectivamente. Se contabilizaron 467 muertes por cáncer de pulmón como causa básica de defunción, y otras 20 citaban dicha enfermedad como causa contribuyente. Para estas 487 muertes se buscó confirmación del diagnóstico a partir del médico que rellenó el certificado de defunción. Como evento de interés se estudió también la mortalidad por otros tumores, enfermedades respiratorias, enfermedades cardiovasculares y otras.

Ejemplo 2: Estudio Prospectivo Europeo de Dieta y Cáncer (EPIC)

El EPIC (European Prospective Investigation into Cancer and Nutrition) es un estudio prospectivo multinacional coordinado desde la Agencia Internacional de Investigación del Cáncer (IARC). Este estudio ha sido diseñado para investigar la relación entre diferentes factores nutricionales y la incidencia de cáncer y otras enfermedades crónicas. Se han reclutado 520.000 participantes en diez países europeos: Alemania, Dinamarca, España, Francia, Grecia, Holanda, Italia, Noruega, Reino Unido y Suecia. El carácter internacional del estudio aumenta la heterogeneidad dietética de los participantes, lo que permite comparar patrones de alimentación muy diversos. El reclutamiento se inició en 1992. Aunque, por lo general, la fuente de reclutamiento de los participantes ha sido la población general, la forma de seleccionar y seguir a los participantes varía de unos países a otros. La información sobre dieta se obtuvo mediante un cuestionario dietético que fue adaptado y validado en todos los países participantes. Se obtuvieron también muestras de sangre, que fueron almacenadas para análisis posteriores. El estudio se plantea un seguimiento mínimo de 10 años. Los casos de cáncer que se van produciendo en la cohorte son detectados mediante los registros de cáncer poblacionales de las áreas del estudio. España contribuye en este estudio con 41.440 participantes, reclutados por los Registros de Cáncer de Granada, Murcia, Oviedo, Pamplona y San Sebastián. El centro coordinador español está en Barcelona.

Ejemplo 3: El estudio de las enfermeras americanas NHS (Nurses' Health Study)

El primer estudio se inició en 1976, y su objetivo principal era investigar los efectos a largo plazo del consumo de anticonceptivos. Se eligieron las enfermeras debido a su alto nivel educativo y a su capacidad para contestar las preguntas relevantes del estudio. El diseño inicial se centró en la población de enfermeras casadas, con edades entre 30 y 55 años en 1976 que vivían en 11 estados de USA, en los que fue posible obtener los listados de las enfermeras y su dirección. Se consideró incluidas en la cohorte a todas las enfermeras que contestaron el cuestionario inicial, (122.000 enfermeras de un total de 170.000 a las que se había enviado dicho cuestionario).

Cada dos años, las enfermeras de la cohorte han de rellenar un cuestionario de seguimiento incluyendo preguntas sobre su salud, consumo de tabaco, uso de anticonceptivos, hormonas y estatus menopáusico.

A partir de 1980 se incluyeron cuestionarios dietéticos con una periodicidad cuatrienal. Además, para tener en cuenta aspectos dietéticos difícilmente investigables a través de cuestionario, se pidió a las enfermeras que enviasen cortes de las uñas de los pies. En 1989 se recogieron muestras de sangre de una submuestra de la cohorte (32.826 mujeres). Finalmente

en 1992 se añadió un cuestionario de calidad de vida, que se cumplimenta también cada cuatro años.

Al NHS-I se ha sumado el NHS-II en 1989, cuyo objetivo es estudiar el efecto de los anticonceptivos orales, la dieta y los estilos de vida en una población más joven, con edades comprendidas entre los 25 y los 42 años al inicio del estudio. Esta generación incluye mujeres que empezaron a tomar anticonceptivos durante la adolescencia y por ello han tenido una mayor exposición a lo largo de la vida. En esta cohorte se han incluido casi 117.000 enfermeras.

Ejemplo 4: La cohorte MACS (Multicentric Aids Cohort Study)

Esta cohorte multicéntrica americana se constituyó para investigar la historia natural de la infección por VIH. EN 1984 se reclutaron 4.954 hombres homosexuales en Baltimore, Chicago, Los Ángeles y Pittsburg, a los que se sumaron otros 668 varones entre 1987-1991, fundamentalmente para representar a las minorías no suficientemente consideradas en la muestra inicial (minorías étnicas y usuarios de drogas parenterales).

En el momento del reclutamiento inicial, se sabía que el SIDA era una enfermedad infecciosa, pero no se había descubierto el agente causal ni se disponía de las técnicas serológicas para su determinación. Cuando dichas técnicas estuvieron disponibles, el análisis de las muestras de sangre almacenadas de forma sistemática cada 6 meses permitió determinar el estatus serológico de los participantes. Se comprobó que el 39% de los participantes eran seropositivos en el momento del reclutamiento (seroprevalentes), mientas que entre los 3.427 pacientes seronegativos, entre 1984 y 1997 se observaron 551 seroconversiones. Esta cohorte ha permitido investigar diferentes aspectos de la historia natural de la enfermedad, incluyendo la prevalencia y la incidencia del VIH, la inmunosupresión, el periodo de incubación, los marcadores de progresión de la enfermedad, la supervivencia desde el diagnóstico de SIDA, los factores de riesgo de la progresión rápida en los seropositivos, la efectividad de las nuevas terapias, etc.

Ejemplo 5: La cohorte de población activa sueca

En los países nórdicos, la existencia de registros poblacionales de gran calidad ha permitido la reconstrucción de cohortes históricas mediante el procedimiento de enlace de registros (record-linkage). Un ejemplo de este tipo de estudios es la cohorte de población activa sueca, identificada a través del censo de población de ese país de 1970, que ha sido seguida un total de 19 años (1971-1989) mediante enlace de registros, utilizando los registros nacionales suecos de mortalidad y de cáncer. La cohorte incluyó únicamente a los trabajadores de 1970 que estuviesen ya presentes en el país, aunque no trabajasen, en 1960. El enlace de registros se realizó utilizando el número de identificación personal, un código de 10 dígitos que identifica de forma inequívoca a cada ciudadano sueco.

La cohorte de trabajadores con edades comprendidas entre los 25 y los 64 años al inicio del seguimiento, incluyó 1.779.646 hombres y 1.066.346 mujeres. En el estudio inicial se presentaba la asociación entre la ocupación desempeñada en 1970 y la incidencia de 30 tipos de cáncer. De forma más específica, se ha investigado la incidencia de cáncer de mama y posteriormente de otras localizaciones.

Frente a otras cohortes históricas similares, la cohorte sueca presenta las siguientes peculiaridades: 1) estudio de ambos sexos y no sólo de los trabajadores varones, 2) cálculo exacto de la contribución de personas-año de cada participante del estudio, y 3) consideración de la ocupación referida en el censo de 1960 para definir subcohortes más específicas para cada

ocupación, incluyendo sólo a los trabajadores clasificados bajo ese código ocupacional en ambos censos.

Ejemplo 6: Cohortes hospitalarias: Estudio de factores pronósticos de cáncer de pulmón no microcítico

Los estudios de factores pronósticos frecuentemente utilizan cohortes hospitalarias de pacientes. El ejemplo que planteamos es un estudio multicéntrico sobre factores pronósticos de pacientes con cáncer de pulmón no microcítico (CPNM) resecable, llevado a cabo de forma prospectiva en 13 hospitales españoles. En el estudio se incluyeron a todos los pacientes con cáncer de pulmón no microcítico, intervenidos en los centros participantes, que cumpliesen los criterios de inclusión: casos con tumor resecable con confirmación histológica, no fallecido como consecuencia directa del procedimiento quirúrgico y sin quimioterapia o radioterapia pre-quirúrgica. En total, el estudio incluyó 465 pacientes.

El objetivo del estudio era evaluar el posible papel como factor pronóstico de determinados biomarcadores: niveles séricos preoperatorios de algunos antígenos relacionados con el cáncer como el antígeno carcinoembrionario (CEA) y CA125, y la sobre-expresión en el tejido tumoral de algunas proteínas como p53 y c-erbB-2. Todas las determinaciones se hicieron de forma centralizada para minimizar la variabilidad debida al laboratorio. Además, en aquellos pacientes que aceptaron participar, se obtuvo información detallada sobre el consumo de tabaco y la historia ocupacional completa. Se investigó también el posible efecto pronóstico del consumo de tabaco, utilizando la información procedente de los últimos 5 años previos al diagnóstico.

El momento de la operación quirúrgica supone el inicio del seguimiento. Los pacientes fueron citados con un intervalo mínimo de 6 meses entre las visitas médicas. Los eventos de interés fueron la muerte por cáncer de pulmón y la aparición de recidiva. También se utilizó el teléfono para localizar a los pacientes que no acudían al hospital y confirmar su estado de salud. Algo original en este estudio ha sido la investigación de la relación entre los factores pronósticos y el tipo de recidiva, diferenciando entre recidiva locorregional y aparición de metástasis a distancia.

2. CLASIFICACIÓN DE LOS ESTUDIOS DE COHORTES

Considerando la referencia temporal del investigador respecto del seguimiento, los estudios de cohortes se clasifican en:

1. Estudios prospectivos: El diseño se realiza a priori, de forma que todo el seguimiento de los miembros de la(s) cohorte(s) se sitúa cronológicamente dentro del periodo del estudio. El estudio de los médicos británicos, el EPIC, el estudio de las enfermeras (NHS) y el estudio de factores pronósticos de CPNM son estudios prospectivos.
2. Estudios retrospectivos o cohortes históricas: El estudio se inicia a posteriori. Se trata de reconstruir el seguimiento de los miembros de la cohorte a través de la información recogida en registros históricos (ej: censos, padrones, registros de cáncer, historias clínicas, etc.). Lógicamente en estos estudios no será posible obtener información que no esté recogida en los registros utilizados. Por ello, el factor limitante es la calidad de las fuentes de información utilizadas. El estudio de la cohorte ocupacional sueca es un ejemplo de cohorte histórica.
3. Estudios ambispectivos: Es una mezcla de los dos anteriores. En el momento de iniciar el estudio, parte del seguimiento de todos o de algunos de los miembros de la cohorte

ya ha transcurrido, por lo que se hace necesario incorporar esta información histórica al resto de los datos recogidos de forma prospectiva. Los estudios de cohortes ocupacionales frecuentemente son ambispectivos. Cuando se decide iniciar el estudio, algunos de los miembros de la cohorte llevan bastante tiempo trabajando en la industria de interés, mientras que los nuevos contratados se incorporan al estudio de forma totalmente prospectiva. El estudio MACS es un ejemplo mixto. Con respecto a la seroconversión, el estudio es prospectivo en los seronegativos y retrospectivo en los positivos.

Por otra parte, en función del grupo de comparación o referencia utilizado se pueden considerar tres tipos de estudios:

1. Estudios con dos cohortes: expuesta y no expuesta: Una vez definida la cohorte expuesta, los investigadores seleccionan e incorporan en el estudio una cohorte no expuesta.
2. Estudios de comparación con la población general: El estudio identifica y sigue a una determinada cohorte de exposición. La frecuencia de aparición de la enfermedad en dicha cohorte es comparada con la observada en la población general. Para ello es necesario disponer de registros poblacionales que suministren dicha información.
3. Comparaciones internas: La cohorte a estudio supone una mezcla de personas expuestas y no expuestas. Las comparaciones se realizan dentro de la propia cohorte.

3. VENTAJAS E INCONVENIENTES DE LOS ESTUDIOS DE COHORTES FRENTE A LOS ESTUDIOS DE CASOS Y CONTROLES

Frente a los estudios caso-control, los estudios de cohortes presentan las siguientes ventajas:

— Permiten obtener información más detallada de los efectos de una determinada exposición. Posibilitan la inclusión en el estudio de uno o más eventos de interés relacionados con dicha exposición.

— Presentan menor propensión a los sesgos de recuerdo y de selección. Por un lado, la relación exposición-enfermedad es estudiada respetando la secuencia temporal. La valoración de la exposición es anterior al evento de interés. Por otro lado, la población origen del estudio está mejor definida y es fácilmente identificable. El sesgo de selección es el punto crítico de los estudios de casos y controles, cuya validez radica en la adecuada inclusión de controles realmente representativos de la población origen de los casos.

— Facilitan el estudio de exposiciones infrecuentes, difíciles de incluir en estudios de casos y controles.

— Favorecen un uso adecuado de las muestras biológicas, que son recogidas antes de que aparezca la enfermedad.

— Posibilitan la medición de la exposición de forma repetida en el tiempo. Con ello permiten valorar los cambios que se producen en la exposición y también calibrar los posibles errores de medición.

— Permiten estudiar la incidencia de la enfermedad en la población a estudio y también el riesgo (incidencia acumulada).

Entre los inconvenientes de los estudios de cohortes, respecto a los casos y controles, figuran los siguientes:

— En general requieren mucho más tiempo. El seguimiento puede durar muchos años, lo que encarece el estudio y dificulta la retención de los participantes. Los estudios de cohortes históricas no sufren dicho inconveniente, pero su calidad está determinada por las fuentes de información utilizadas.
— Cuando el evento de interés es infrecuente (enfermedades de baja incidencia) es necesario incluir en el estudio un número elevado de participantes.
— Los estudios de cohortes basadas en grupos específicos de exposición, como pueden ser determinados sectores ocupacionales, no permiten estimar directamente la fracción de enfermedad atribuible a la exposición. Para ello, es necesario conocer cual es la frecuencia de exposición en la población general.

Vistos los pros y contras, podemos concluir que los estudios de cohortes y los caso-control son dos herramientas complementarias. Los estudios de cohortes proporcionan una población bien definida en la que identificar los casos de forma no sesgada y con información disponible de la secuencia temporal del seguimiento. Los caso-control concentran el esfuerzo en los "individuos informativos". Por eso muchos estudios de cohortes incorporan la estrategia de utilizar diseños de caso-control dentro de la cohorte.

4. ASPECTOS BÁSICOS DEL DISEÑO DE LOS ESTUDIOS DE COHORTES

A la hora de diseñar un estudio de cohortes es importante definir de forma clara los siguientes aspectos:

— Selección de la cohorte: Criterios de inclusión
— Fecha de entrada y de salida de los participantes en el estudio
— Seguimiento
— Eventos finales de interés
— La exposición
— Factores de confusión
— Confidencialidad
— Estrategias de análisis
— Potencia estadística
— Optimización de recursos: estudios de caso-control sobre una cohorte

5. SELECCIÓN DE LA COHORTE

La selección de la(s) cohorte(s) viene motivada por la naturaleza de la exposición, las características de la enfermedad estudiada y la posibilidad de seguimiento del grupo elegido.

Respecto a la exposición, es importante que la cohorte elegida incluya personas con un grado de exposición suficiente como para desarrollar el/los eventos de interés. En este sentido, el NHS-II incluye un colectivo nuevo de enfermeras en las que la posible exposición a hormonas contraceptivas ha sido mucho mayor que la de la cohorte inicial. Por otra parte, si el estudio pretende establecer comparaciones internas dentro de la cohorte, ésta debe ser suficientemente heterogénea en cuanto a niveles de exposición. Así, el carácter multinacional del estudio EPIC permitirá investigar la influencia de patrones dietéticos que normalmente presentan poca variabilidad en una región o incluso en un país.

Respecto al evento o enfermedad de interés, cohortes como el MACS fueron diseñadas para investigar los factores de riesgo de enfermedades de origen desconocido. En este sentido, interesa incluir en la(s) cohorte(s) colectivos con una elevada incidencia. Por este motivo, el reclutamiento inicial en el estudio MACS incluyó 5.000 varones homosexuales, colectivo que sufrió el primer impacto de la epidemia de SIDA en Estados Unidos.

Teniendo en cuenta el seguimiento, es importante que el colectivo sea fácil de seguir a lo largo del estudio y que el porcentaje de abandonos del estudio (pérdidas esperadas) sea bajo. Ello ha motivado la creación de cohortes de médicos y de enfermeras para investigar exposiciones no relacionadas directamente con su trabajo, como son el consumo de tabaco y de anticonceptivos orales. Estos colectivos tienen un mayor grado de motivación y están, además, altamente cualificados a la hora de proporcionar información sobre la exposición e incluso sobre los eventos de interés.

Finalmente, la principal ventaja de cohortes históricas como la cohorte de población activa sueca es su carácter poblacional, lo que elimina los sesgos de selección y permite estimar la incidencia de cáncer y el riesgo relativo en ocupaciones de baja frecuencia.

La base poblacional de un estudio de cohortes se define como la experiencia de personas-tiempo incluida en ese estudio concreto. Los resultados del mismo se refieren precisamente a esa base poblacional. Para que dicha base poblacional sea fácilmente identificable, las reglas de inclusión y exclusión deben ser claras.

6. FECHAS DE ENTRADA Y SALIDA: CONTRIBUCIÓN DE CADA MIEMBRO DE LA COHORTE AL ESTUDIO

La fecha de entrada supone el inicio de la contribución de cada miembro de la cohorte al estudio, en términos de personas año. La fecha de entrada puede, y de hecho suele, no coincidir con el inicio de la exposición, ni con la fecha en la que dicha persona ha completado lo que podríamos considerar como “dosis mínima” de exposición relevante.

La fecha de salida está condicionada por tres factores: 1) el desarrollo del evento o enfermedad de interés, en caso de que dicho evento sea único; 2) el cierre del estudio, y 3) el abandono del estudio por la causa que sea.

A la hora de establecer el nivel de exposición, es importante considerar la información anterior a la inclusión del sujeto en el estudio. Sin embargo, ese periodo no debe contribuir en el cómputo de personas año. Dicha inclusión supondría un sesgo de selección, ya que el diseño del estudio ha dejado fuera a aquellos miembros de la población objeto del estudio que ya no están presentes (no son reclutables) al inicio del mismo, ya sea por haber fallecido, haber desarrollado la enfermedad previamente o haber abandonado dicha población.

En la tabla 6-1 se presenta la información sobre la entrada y salida y el tiempo biológicamente relevante en cada uno de los ejemplos citados al inicio de este capítulo.

Tabla 6-1: Periodo biológicamente relevante y periodo de estudio en los estudios de cohortes citados.

Estudio	Periodo biológicamente relevante (PBR)	Entrada en el estudio	¿Incluye el estudio todo el PBR?
Médicos británicos	Del inicio del consumo de tabaco a la muerte por cáncer de pulmón u otras causas	Primer cuestionario	No
EPIC	Del inicio de consumo de determinado nutriente a la aparición de un det. tumor	Primer cuestionario y toma de muestras biológicas	No
NHS	Del inicio del consumo de anticonceptivos a la aparición de uno de los eventos de interés	Primer cuestionario	No
MACS	Seroconversión - desarrollo de SIDA y posteriormente muerte	Primer cuestionario y muestra de sangre	Sólo en sujetos seronegativos
Cohorte sueca	Del inicio de una determinada ocupación a la aparición de un determinado tumor	1-Enero-1971	No
Factores pronósticos CPNM	Cirugía - recidiva tumoral – muerte	Fecha cirugía	Si

7. SEGUIMIENTO

La calidad del seguimiento es el principal factor para definir la calidad de un estudio de cohortes. En general, los objetivos del seguimiento son tres:

— Identificar la aparición del evento.
— Identificar las pérdidas del estudio y el momento en que se producen.
— Identificar cambios relevantes en la exposición y, si es necesario, en otros factores de riesgo relevantes.

El tipo de seguimiento es diferente según los estudios. En algunos, el seguimiento es pasivo, basado exclusivamente en la información disponible en los registros poblacionales, como el caso de los médicos británicos o la cohorte sueca. En otros, los sujetos participantes deben acudir de forma periódica al hospital o centro correspondiente, como es el caso del MACS y del estudio de factores pronósticos de CPNM. En otros casos, el seguimiento se realiza por correo, a través de encuestas (ejemplo: NHS).

La periodicidad del seguimiento viene determinada por la dinámica de la enfermedad y la necesidad de mantener a los participantes dentro del estudio. Aunque las técnicas de análisis estadístico permiten considerar la información parcial proporcionada por las personas que abandonan el estudio, un alto porcentaje de abandonos o pérdidas puede comprometer seriamente la generabilidad de los resultados. Por ello, muchos estudios de cohortes han desarrollado estrategias para mantener la tasa de participación de los sujetos incluidos (ver tabla 6-2).

Tabla 6-2: Estrategias para maximizar la adherencia.

En el reclutamiento: - Investigar la intención de permanecer en la cohorte - Hacer que el participante complete con éxito las tareas que se requerirán de él en el seguimiento - Informar ampliamente de los requerimientos de su participación en el estudio - Recoger información para el seguimiento: dirección, teléfono, número SS, lugar y fecha de nacimiento, etc - Recoger información similar de contactos
Identificación con el estudio: - Crear un logo y un lema - Mandar cartas, felicitaciones, información sobre la marcha del estudio, etc.
Frecuencia de contacto: - Contactos regulares cada 6-24 meses - Recoger como mínimo información sobre el evento - Diseñar un sistema de rastreo
Responsables del seguimiento: - Entrenados y entusiastas - Que faciliten la comunicación - Que sean flexibles en las citación de los participantes
Uso de incentivos - Distribuir entre los participantes objetos con el logotipo del estudio - Cubrir los gastos de los envíos por correo - Informar periódicamente a los participantes de los resultados del estudio

Es importante que el seguimiento cuente con un dispositivo capaz de detectar los abandonos del estudio y el momento en que se producen. En situaciones de seguimiento pasivo, cuando el diagnóstico de la enfermedad se realiza a través de registros, el investigador puede estar considerando dentro del estudio a sujetos que de hecho están ya fuera de él, lo que se traduce en una subestimación de la incidencia real. Por ejemplo, la cohorte de población activa sueca fue seguida mediante los registros suecos de cáncer y de defunción. Todas las personas que refirieron una ocupación concreta en 1970 se consideraron dentro del estudio mientras no figurase su fecha de defunción en el registro de mortalidad. Sin embargo, algunos trabajadores han podido salir del país y por ello abandonar el estudio. En este caso concreto, la imprecisión cometida al considerar a estas personas dentro de la cohorte es pequeña, dado que la tasa de emigración en Suecia en los años 70 y 80 fue menor del 1 por mil.

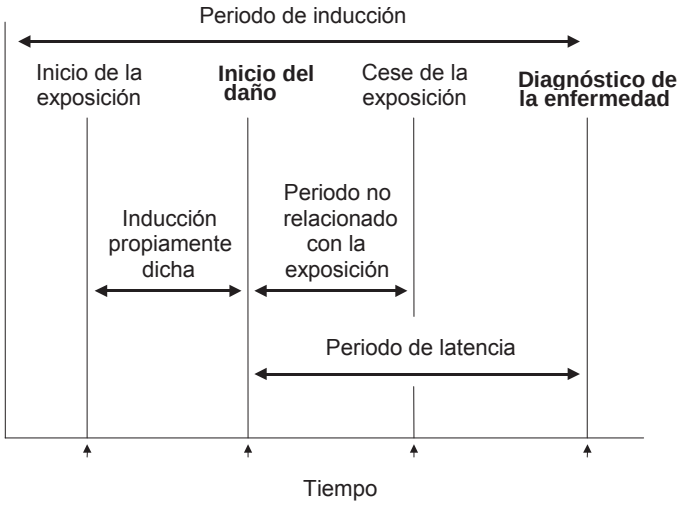
8. EVENTOS FINALES DE INTERÉS

Durante el seguimiento es necesario disponer de un sistema que identifique en qué momento se produce el evento o enfermedad a estudio. En estudios de mortalidad, como la cohorte de médicos británicos, el momento en que se produce el evento es fácil de constatar. Lo que resulta más problemático en estos estudios es considerar la causa de defunción. La calidad de los certificados de defunción es heterogénea según el tipo de enfermedad considerada. En el caso de la mortalidad por cáncer, diferentes estudios han mostrado que la información contenida en los certificados de defunción es razonablemente válida.

En los estudios de incidencia basados en registros poblacionales, la fecha del evento es la fecha de diagnóstico que figura en el registro. En el estudio MACS, como fecha de seroconversión se consideró la media entre la última visita con serología negativa y la primera con serología positiva. En el estudio de factores pronósticos de CPNM, se consideró como fecha de recidiva la primera visita o exploración positiva.

La mayor parte de enfermedades crónicas plantean la necesidad de periodos extensos de seguimiento, debido a la magnitud de los periodos de inducción y latencia. Aunque algunos autores usan ambos términos como sinónimos, el periodo entre la exposición y la aparición de los síntomas de la enfermedad puede dividirse en dos partes (ver figura 6-1). El periodo de inducción es el intervalo de tiempo necesario para que el agente causal (sólo o con la concurrencia de otros factores) sea suficiente para iniciar el daño. A partir de este punto comenzaría el periodo de latencia, durante el cual la enfermedad progresa hasta el momento en el que es diagnosticable. Este periodo se reduce, por ejemplo, con programas de screening que detectan la enfermedad presintomáticamente. Si, una vez que se inicia el proceso patológico, la enfermedad puede revertir sin haber sido diagnosticada, ambos periodos se solapan; si éste no es el caso, el periodo de inducción y el de latencia son consecutivos. Tras el inicio de la enfermedad, la exposición puede seguir existiendo: en algunos casos la presencia del agente ya no tiene efecto sobre la evolución de la misma, mientras que en otros puede potenciar su progreso (agente promotor).

Figura 6-1: Relación temporal entre exposición y enfermedad.



Actualmente se investiga la posibilidad de utilizar biomarcadores como eventos intermedios o marcadores de efecto precoz, que permitirían disminuir la duración del estudio e instaurar medidas preventivas antes de que la enfermedad avance (ver figura 6-4). Por ejemplo, en el estudio de la historia natural de infección por VIH, la caída de linfocitos CD4+ es un determinante de progresión de enfermedad. Para las enfermedades cardiovasculares se han utilizado los niveles de colesterol y de lipoproteínas, como eventos intermedios. En relación al cáncer, los marcadores de efecto precoz utilizados incluyen cambios genéticos en las células, como son la aparición de aberraciones cromosómicas y micronúcleos.

9. VALORACIÓN DE LA EXPOSICIÓN

La medida de la exposición es un punto crucial en cualquier estudio epidemiológico. Al hablar de exposición, sin embargo, no nos referimos sólo al posible contacto con un agente de interés, sino también a todos aquellos factores relacionados con el sujeto que pueden influir en la relación entre el agente y la enfermedad. Una adecuada recogida de información sobre hábitos de vida, indicadores de situación socioeconómica, problemas de salud u otras características que varíen la susceptibilidad personal a los agentes de interés y puedan modificar el efecto que se desea analizar, es casi tan importante como la misma valoración de la exposición al factor que motiva la realización del estudio.

En general, los estudios de cohortes son un diseño muy atractivo para estimar la exposición ya que garantizan que la medición se realiza antes de que se produzca la enfermedad. Así se cumple uno de los principios básicos en causalidad: "la causa precede al efecto". Esta secuencia temporal tiene otras ventajas añadidas:

1. La medida de la exposición no está condicionada por la enfermedad ni por los tratamientos que se puedan aplicar para ella. Este aspecto es muy importante si se usan marcadores biológicos para estimar la exposición.
2. Las variables de confusión (por ejemplo, aspectos relacionados con la dieta, ejercicio o hábitos de vida) se miden también antes de la aparición de la enfermedad. Así se evitan tanto el sesgo de recuerdo, como el posible efecto que el diagnóstico de la enfermedad puede tener sobre las respuestas y conductas de los sujetos. Como contrapartida, sin embargo, hay que señalar que lo habitual es que la información sobre variables asociadas se recoja con menos detalle que en un estudio de casos-controles, en el que se puede focalizar más la atención hacia los aspectos relacionados con el evento de interés. Existe, además, otro problema derivado del momento de recogida de los datos: no podremos tener en cuenta nuevos factores de riesgo que se identifiquen posteriormente, porque no dispondremos de información sobre los mismos. La modificación del instrumento de medida (e.j cambios en el cuestionario) puede permitir tener dicha información a partir de un momento dado. Por ejemplo, el "Nurses' Health Study" incorporó un cuestionario dietético a partir del cuarto año de seguimiento.

Aunque muchas veces es difícil conocer la fecha de la exposición inicial, el diseño de la cohorte permite disponer al menos de una estimación de la exposición en un momento concreto para todos sus miembros. Si el seguimiento de la cohorte es largo, dispondremos de un rango muy amplio de tiempos diferentes entre el momento de la medición y el momento de aparición del proceso de interés. Esto puede permitir investigar hipótesis relacionadas con diferentes periodos de latencia entre exposición y enfermedad.

10. LA ESTIMACIÓN DE LA EXPOSICIÓN

Un estudio de cohortes es esencialmente un estudio analítico; la estimación de la exposición, por tanto, se realiza con un fin: estudiar la asociación entre aquélla y unos efectos determinados. Esto hace que el investigador tenga que plantearse a priori la relación entre ambas y cuáles son los aspectos a tener en cuenta para representar la "dosis" de exposición relevante en el contexto de esa enfermedad. Para entender a que nos referimos vamos a usar como ejemplo el estudio del efecto a largo plazo de una dosis excesiva de radiaciones ionizantes ocurrida accidentalmente en trabajadores de una central nuclear. La estrategia de medida será diferente en función de las asunciones previas, por ejemplo: a) si presumo que sólo se producirán efectos dañinos en caso de que la radiación haya superado un determinado dintel, b) si lo que presupongo es que todos los trabajadores pueden sufrir consecuencias adversas por este accidente y que a mayor exposición el riego de alteraciones es mayor, o c) si considero que el efecto de esta exposición va a estar determinado también por la exposición a radiación que el sujeto acumule a lo largo de su vida. Si precisamente lo que se pretende es caracterizar esta relación, es necesario tener presentes todas estas posibilidades y recoger la información necesaria.

Tras haber pensado cuál es la mejor manera de aproximarse a esa teórica medida de "dosis" ideal, el investigador tiene que identificar qué exposiciones incluyen el agente de interés, definir los ítems que necesita recoger para obtener información sobre la duración, frecuencia e intensidad de la exposición en el periodo específico de riesgo que se ha determinado, y desarrollar el algoritmo que combina dichas variables para llegar al constructo de dosis elegido.

Aunque en este capítulo vamos a tratar dosis y exposición prácticamente como si ambos términos fuesen sinónimos, hay que señalar que muchos autores diferencian claramente entre exposición, referida a medidas tomadas en el ambiente, es decir fuera del sujeto, y dosis, que se mide bien en tejidos humanos o en el punto de contacto entre sujeto y medio. Ott, en la Enciclopedia de Salud y Seguridad en el Trabajo de la OIT, define la exposición como la concentración de un agente en la frontera entre individuo y medioambiente, y la dosis como el depósito o absorción de un agente por unidad de tiempo, y se estima combinando datos de mediciones ambientales con hipótesis tipificadas o con medidas directas. La diferencia entre las dos depende de características fisiológicas, de los patrones de actividad de los sujetos y de las variaciones en la concentración de los agentes en el tiempo y el espacio. En general se considera que las medidas externas son más útiles que las internas cuando las exposiciones están muy extendidas y las diferencias entre grupos son más importantes que las interindividuales.

10.1 ¿Cómo medimos la exposición? Exposición y tiempo

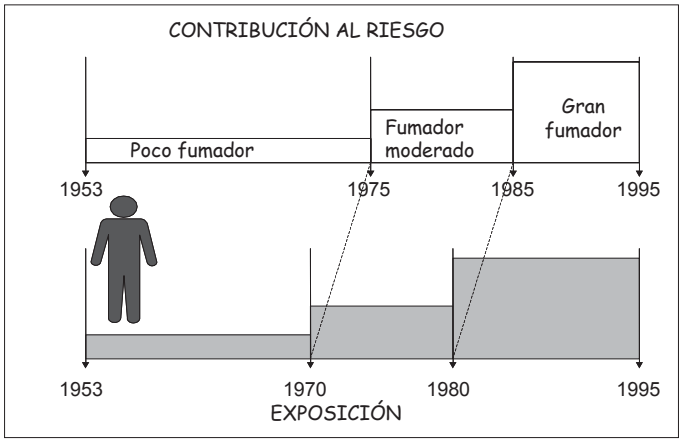
En el estudio de cohortes más sencillo, la exposición sería una condición permanente y fácilmente identificable, por lo que asignar individuos al grupo de expuestos sería prácticamente una tarea rutinaria. Sin embargo, en la realidad las exposiciones de interés rara vez son constantes y a menudo son difíciles de medir. Otros factores de riesgo presentan también este problema: existen características que no cambian en el tiempo, como el sexo, pero otras pueden variar durante el seguimiento, como por ejemplo la ocupación o la residencia habitual.

El primer aspecto a definir es el periodo de exposición que consideramos relevante: ¿una medida de la exposición actual es suficiente o se necesita una medida pasada? Casi todas las medidas de laboratorio ("objetivas") son actuales, aunque algunas, como las derivadas de rasgos heredados, son fijas. Sin embargo, a menudo es necesario tener en cuenta la posible existencia de periodos de inducción y de latencia, es decir, los periodos de tiempo que se consideran necesarios para que el agente pueda producir el daño (inducción) y para que éste sea detectable (latencia). En la figura 6-1 se muestra un esquema de la posible relación entre un tóxico y una enfermedad. Si estamos estudiando, por ejemplo, tabaquismo y cáncer de pulmón en mujeres, debemos suponer que ha de pasar algún tiempo desde que la mujer empezó a fumar para que el tabaco pueda generar daño y se origine el cáncer. Al principio, además, el tumor que empieza a desarrollarse en el pulmón es prácticamente indetectable, y ha de pasar un tiempo hasta que produzca síntomas y sea diagnosticado.

Otra pregunta importante es durante que periodo se recogerá la medida de la exposición: ¿es suficiente con una única medida, necesitamos medidas repetidas para aumentar la precisión o es preferible actualizar la información con medidas repetidas del marcador en diferentes momentos desde la medida inicial? De nuevo es fundamental el esquema teórico del investigador para determinar cuál es la estrategia adecuada. En muchos análisis epidemiológicos se usa una medida resumen, asumiendo implícitamente que es éste el único aspecto de la exposición que determina el riesgo actual. Las medidas resumen más habituales son el nivel de exposición en el momento de la medida, la exposición media o mediana si existen medidas repetidas, la exposición máxima registrada o la exposición acumulada, es decir la suma de exposición en cada categoría por el tiempo que el individuo pasa en la misma. El seguimiento en el tiempo de los individuos, sin embargo, conlleva el problema añadido de clasificar la "experiencia" de un individuo en diferentes categorías de exposición en distintos momentos del tiempo. Por ejemplo: en la figura 6-2 podemos ver cómo se podría categorizar la exposición acumulada al tabaco de un sujeto, que podría pertenecer a la cohorte de médicos británicos previamente mencionada, y que fuma sin interrupción entre 1953 y 1995. Inicialmente el sujeto fuma entre uno y 5 cigarrillos al día, y hasta 1970 no acumula las 1000 cajetillas-vida que hemos definido como límite para

ser fumador moderado. Si, además, tenemos en cuenta un posible periodo de latencia de 5 años, consideraríamos que las personas-año que aporta este sujeto entre 1953 y 1975 corresponden al estrato de poco fumador. Nuestro individuo incrementa algo su consumo de cigarros y en solo 10 años alcanza las 2000 cajetillas-vida que determinan el paso a gran fumador. Las personas-año correspondientes al periodo 1975-85 corresponden a fumador moderado y, a partir de aquí, a gran fumador hasta que el sujeto muera, se pierda o se abandone el seguimiento.

Figura 6-2: Categorización de la exposición acumulada al tabaco.



10.2 Herramientas para estimar la exposición

La calidad en la medida de la exposición determinará en muchos casos la validez de los resultados del estudio. Por este motivo deben usarse instrumentos de medida correctamente validados, y es conveniente probarlos en una pequeña muestra, revisarlos y probarlos de nuevo en un estudio piloto. En la tabla 6-3 se recogen las posibles fuentes de error que hay que tener presentes en la estimación de la exposición. La precisión y el grado de detalle con el que se recoge la información es otro aspecto importante. En general hay que intentar, en la medida de lo posible, cuantificar. La información cuantitativa permite estudiar la relación dosis-respuesta y la dosis de exposición mínima necesaria. Es necesario recoger además toda la información sobre las condiciones en las que se produce la exposición si se cree que pueden modificar la misma.

Tabla 6-3: Posibles fuentes de error en la medida de la exposición

- Error en el instrumento de medida o en su protocolo
  - Falta de exhaustividad en cubrir todas las fuentes de exposición
  - Inclusión de exposiciones que no tienen el agente activo
  - Periodo evaluado distinto del periodo relevante
  - Preguntas ambiguas o mal entendidas.
- Errores en el protocolo de uso del instrumento
  - Poca especificación de las instrucciones
  - Imprevisión sobre la forma de manejar situaciones y respuestas no previstas
- Cambios en la precisión a lo largo del tiempo
  - Cambio del instrumento de medida
  - Cansancio del entrevistador
  - Ausencia de estandarización del procedimiento
  - Degradación de muestras biológicas
- Errores en el procesado de los datos
  - Errores en la grabación
  - Errores en la conversión de la exposición al agente activo
  - Variable indicadora o resumen de la exposición inadecuada

Teniendo presentes todos estos aspectos, el investigador debe elegir un instrumento de medida atendiendo a dos puntos básicos: el factor a estudiar y la factibilidad de su uso en el estudio concreto que se plantea. Entre las formas más habituales de medir la exposición podríamos citar las siguientes:

— Entrevistas, cuestionarios o, con menor frecuencia, diarios estructurados. Son fundamentales sobre todo a la hora de disponer información relativa a hábitos o conductas. Permiten tener información cuando no existen registros objetivos, aunque pueden estar sujetos a errores difíciles de controlar. Por ejemplo, las contestaciones a determinadas preguntas pueden depender de la consideración social del comportamiento sobre el que se pregunta.
— Medidas en el ambiente del sujeto, bien procedentes de registros existentes o bien obtenidas expresamente para el estudio (p. ej, registros de contaminación ambiental en una ciudad).
— Medidas en el entorno inmediato del individuo (por ejemplo, monitores ambientales en una estancia determinada de una empresa).
— Dosis individual (por ejemplo, monitorización personal).
— Medidas de concentración en tejidos humanos o productos metabólicos (biomarcadores).
— Marcadores de efectos fisiológicos (biomarcadores de dosis efectiva).
— Muchas veces, en los estudios se combinan varios instrumentos de medida, como se puede comprobar en las cohortes que se han citado como ejemplo. Cada vez es más habitual incorporar recogida de materiales biológicos (sangre, DNA...), acompañados de una encuesta, que resulta imprescindible para tener información sobre otros factores de riesgo o posibles factores de confusión. En la tabla 6-4 se resumen las estrategias que se han utilizado para medir la exposición en las cohortes que se están usando como ejemplo.

Un punto elemental a tener en cuenta si hablamos de cohortes grandes es el coste en dinero y en esfuerzo. Entre las estrategias posibles para realizar estudios de cohortes con costes reducidos podríamos destacar las siguientes:

Los estudios *basados en enlaces de registros*, como la cohorte de población activa sueca, en los que bases de datos que contienen información sobre exposición se asocian a través de números de identificación personales con registros que contienen la información sobre enfermedad. Es bastante habitual y más económico recurrir a registros preexistentes (listados de ocupaciones, etc...), aunque muchas veces esto hace que se disponga de poca información sobre posibles variables de confusión

Los estudios que reclutan la cohorte y recogen la información sobre exposición a través de *cuestionarios que se envían por correo,* que han de ser necesariamente breves y concisos. Esta es la estrategia que se ha usado, por ejemplo, en el "Nurses' Health Study". En este caso, además, se solicitó a los mismos miembros de la cohorte el envío de muestras biológicas (uñas).

Los que *miden la exposición sólo en una muestra de la cohorte.* Usan diseños que permiten reducir costes, minimizando la pérdida de información, como veremos más adelante.

Tabla 6-4: Instrumento de medida de la exposición y periodicidad de la misma en algunos de los estudios de cohortes citados como ejemplo.

Estudio	Instrumento de medida de exposición	Periodicidad de la medida de la exposición	Algunos aspectos interesantes del estudio
EPIC	Cuestionarios postales o en entrevista Examen físico para medidas antropométricas Muestras biológicas: sangre	Cuestionario reducido a los 3-4 años del reclutamiento	El estudio incluyó una exhaustiva investigación metodológica para garantizar la validez de los cuestionarios y medidas utilizados que ha sido publicada independientemente.
NHS-I	Cuestionarios postales	Cada 2 años la información sobre uso hormonal, estatus menopausico y tabaco. Cada 4 años información sobre dieta y calidad de vida	Ha ido incorporando cuestionarios sobre nuevos aspectos a lo largo del estudio (dieta en 1980, calidad de vida en 1992)
	Muestras biológicas: Recogida de uñas de los dedos de los pies a los 6 años del inicio (68.000 mujeres) Muestra de sangre domiciliaria a los 13 años del inicio (33.000 mujeres aprox.)	Toma única Prevén repetir la muestra en el año 2000	Algunas tomas de muestras han sido realizadas por los mismos miembros de la cohorte
MACS	Cuestionario autoadministrado Entrevista personal Examen clínico y de laboratorio	Se repite todo cada 6 meses	
Cohorte sueca	Enlace de registros de Cáncer, Mortalidad y censos de 1970 y 1960		Se puede combinar con matrices de ocupación-exposición expresamente diseñadas para esta cohorte
Factores pronósticos CPNM	Encuesta Muestras biológicas	Medida única al inicio del estudio	

10.3 Algunos instrumentos de medida interesantes

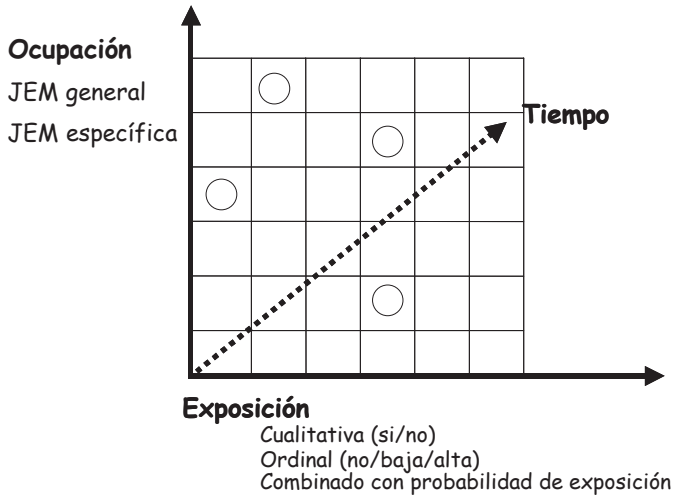
Matrices de Ocupación-exposición

El entorno laboral es una fuente muy rica para el estudio de tóxicos en humanos. En una ocupación, sin embargo, suele haber exposición a múltiples substancias; y por otra parte existen muchas ocupaciones que tienen exposiciones comunes. En muchos casos, la información relevante es la exposición a la sustancia de interés y no la ocupación desempeñada.

Una matriz de ocupación-exposición (Job-Exposition Matrix: JEM) (figura 6-3) es una herramienta para convertir las ocupaciones en exposiciones potenciales. En su forma más sencilla, los puestos de trabajo se clasifican en expuestos o no expuestos. La estimación de la exposición en un trabajo concreto puede basarse en medidas directas o en opiniones de expertos basadas en la descripción de los puestos de trabajo, y será más acorde con la exposición real cuanto más específica sea la definición del mismo. Si definimos categorías ocupacionales (p.ej. empleados de tiendas),

el error en la estimación será superior al que se tendrá si se estudia cada ocupación de forma independiente en cada sector industrial (p.ej. empleadas de tiendas en gasolineras y empleados de tiendas en un centro comercial). Aún así, no todos los sujetos pertenecientes a una misma combinación ocupación-industria están expuestos a las substancias de interés. A veces, la clasificación de la exposición se basa en la proporción estimada de personas en ese empleo que soportan niveles de exposición por encima de un nivel determinado. No obstante, las JEMs no recogen toda la variabilidad que puede existir entre los trabajadores de una misma ocupación, ni la debida a factores personales, por lo que siempre habrá errores de clasificación.

Figura 6-3: Esquema de una matriz de ocupación-exposición (JEM).



En algunos países, como Suecia, se han elaborado JEMs generales de exposición a químicos o a campos electromagnéticos, que se han aplicado a la cohorte de población activa citada como ejemplo. De esta manera se consigue estudiar la exposición a estos agentes en cohortes muy grandes (nacionales) de una forma barata y eficiente.

Biomarcadores

Un biomarcador de exposición (figura 6-4) es un estimador de dosis de un determinado agente medido en muestras o sistemas biológicos. Puede ser un compuesto exógeno al individuo, un metabolito de aquél o un producto interactivo exo-endógeno, como las combinaciones de tóxicos con proteínas del sujeto. Resultan especialmente útiles si hay varias vías posibles de exposición (p.ej, ingesta y absorción dérmica), cuando se usan medios de protección o en condiciones de exposición imprevisibles. Otros biomarcadores relevantes para los estudios de cohortes son los marcadores de susceptibilidad genética: diferencias en la velocidad de acetilación hepática, por ejemplo, hacen que varíe la probabilidad de padecer una enfermedad concreta tras la exposición a un tóxico que se inactiva por esta vía. Modifican por tanto el efecto de la exposición (interacción).

Figura 6-4: Tipos de biomarcadores.

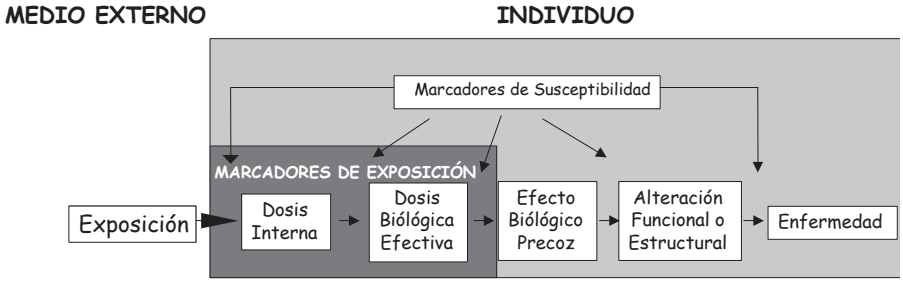


Tabla 6-5: Biomarcadores en estudios de cohortes: ventajas y desventajas

Ventajas del uso de biomarcadores:

1. Puede permitir la integración de varias vías de exposición.
2. No hay necesidad de estimar el efecto de las medidas de protección.
3. Aunque aparentemente dos sujetos estén expuestos a concentraciones similares de una sustancia, en realidad pueden existir diferencias individuales de "ingesta"
   - Por diferencias en exposiciones medioambientales no conocidas.
   - Por diferencias en la biotransformación de la sustancia.
4. Los biomarcadores pueden dar una idea más real de la exposición concreta de un individuo.
5. Si varios agentes tienen un metabolito común puedo estudiar su efecto conjunto.
6. Si la variabilidad intra e inter-individuos es menor que la del medio, puede ser una estrategia muy eficiente de medición de la exposición.

Desventajas del uso de biomarcadores

1. No siempre hay buena correlación con la exposición ambiental. Es muy importante por tanto validar previamente el marcador.
2. Cuidado con el periodo al que se refieren: no siempre encontraremos biomarcadores que se correlacionen con el momento de exposición que queremos estudiar.
3. Los procesos toxicocinéticos y de eliminación modifican las concentraciones de las sustancias de interés y pueden verse alterados por muchas causas (p.ej. exposición simultanea a otro agente no controlado)
4. Si las diferencias interindividuales por variaciones en la entrada, biotransformación y excreción del tóxico son mayores que las medioambientales, pueden ser mejores las segundas
5. Muchas veces no son específicos de un agente concreto: hay que descartar factores de confusión o exposiciones combinadas.
6. Debemos disponer de niveles base o de referencia para interpretarlos y conocer muy bien sus posibles fuentes de variabilidad
7. Puede que no toda la cohorte esté dispuesta a proporcionar muestras, con lo que tendríamos poca potencia. Esto es más evidente para las medidas repetidas, sobre todo si la muestra es invasiva.

Los biomarcadores presentan un innegable atractivo para los investigadores; sin embargo, como todos los instrumentos de medida, han de ser específicos, fiables y precisos. Sus ventajas y desventajas se resumen en la tabla 6-5. La adecuada selección del biomarcador debe tener en cuenta: a) con qué periodo de exposición se relaciona el biomarcador seleccionado, b) cómo afectan los factores metabólicos u otros factores personales del huésped a los niveles del biomarcador, y c) las características de la muestra que se precisa (invasiva o no, cantidad de sustrato, posibilidad de degradación en el tiempo, coste, posibilidad de tomas seriadas si es preciso...), ya que determinan la logística del estudio.

11. OTROS FACTORES DE RIESGO: CONFUSIÓN E INTERACCIÓN

La necesidad de obtener información sobre exposición a otros factores viene motivada por dos razones: 1) la naturaleza multicausal de la etiología de la mayor parte de las enfermedades, 2) los factores que determinan la presencia o ausencia de la exposición de interés.

El estudio de cohortes es de carácter observacional, es decir, el investigador estudia la exposición en aquellos grupos de población en los que ésta se produce. Normalmente, la mayor parte de los factores de riesgo no se distribuyen de forma aleatoria en la población, por lo que con frecuencia los determinantes de la enfermedad están asociados, positiva o negativamente, con la exposición bajo estudio. Esta asociación hace necesario identificar y recoger información sobre los posibles factores de confusión para controlar su efecto en el análisis.

Además, en el estudio de la dinámica de la enfermedad es interesante constatar si existen sinergias entre la exposición bajo estudio y la presencia de otros factores de riesgo (presencia de interacciones).

12. CONFIDENCIALIDAD

El establecimiento de una cohorte implica prever un seguimiento para los individuos que la componen. Esto hace imprescindible que exista un identificador para cada sujeto, que permita asignarle la información sobre su exposición o su estatus vital. Sin embargo, como en el resto de los estudios epidemiológicos, asegurar la confidencialidad de la información es un requisito fundamental para cumplir con los principios éticos exigibles a toda investigación. En España, el tratamiento de los datos personales se rige por la ley Orgánica 15/1999 de Protección de Datos de Carácter Personal, que establece las garantías que deben observarse durante la obtención, custodia y cesión, si es el caso, de los mismos. En esta norma, los datos relativos a la salud son protegidos especialmente, con restricciones y controles más estrictos que con otro tipo de informaciones. En el caso de incluir datos sobre enfermedades, es habitual tener que recurrir a las historias clínicas de los sujetos, siendo entonces también de aplicación la ley 41/2002, reguladora de la Autonomía del Paciente y de los Derechos y Obligaciones en Materia de Información y Documentación Sanitaria y la Ley 14/1986, General de Sanidad, que permiten el acceso a la información con fines epidemiológicos o de investigación, aunque explicitan que, como regla general, dicho acceso debe realizarse de forma que quede asegurado el anonimato. En cualquier caso, es responsabilidad del investigador garantizar la confidencialidad de la información que está a su cargo.

13. ESTRATEGIAS DE ANÁLISIS

El análisis de los estudios de cohortes ha de tener en cuenta la dimensión temporal del estudio. La existencia de pérdidas y la diferente duración del seguimiento de los distintos miembros de la cohorte implican la necesidad de técnicas analíticas especiales que, a la vez, permitan considerar toda la información aportada por los integrantes de la cohorte. Las estrategias de análisis más usadas “desintegran” el tiempo de observación de cada uno de los participantes. La medida de frecuencia más adecuada es la tasa de incidencia. La unidad de análisis ya no es el individuo, sino la unidad temporal de observación. Existen dos estrategias diferentes de análisis en los estudios de cohortes: el análisis supervivencia y el análisis personas-tiempo, pero ambas requieren disponer de:

— Información precisa sobre el inicio y el final del seguimiento de cada participante.
— Definición de una escala temporal adecuada (edad, año calendario, tiempo tras la intervención, etc).
— Estatus del participante en el momento de salida del estudio (desarrollo o no de la enfermedad).

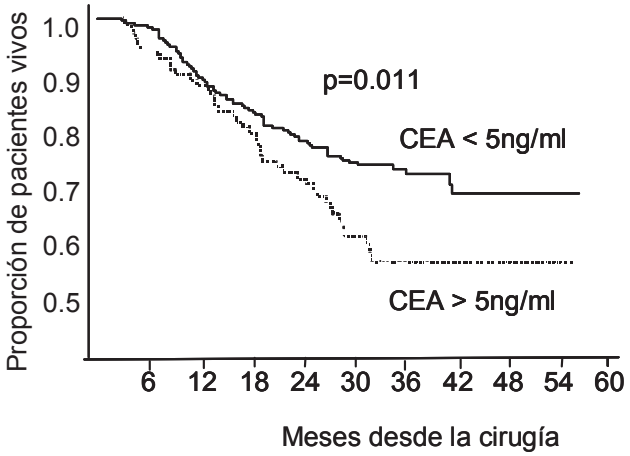
Es importante tener en cuenta, que en el análisis, todos los individuos con información incompleta, es decir, aquellos que no desarrollan el evento de interés, son tratados de la misma manera. Conceptualmente, sin embargo, cabe establecer diferencias entre los participantes que están presentes en el momento del cierre del estudio y aquellos que han abandonado el mismo a lo largo del seguimiento. Las técnicas de análisis consideran que las pérdidas son aleatorias, pero las personas que abandonan el estudio pueden hacerlo por razones o factores relacionados con la enfermedad estudiada o con la exposición de interés. También es importante tener en cuenta la mortalidad competitiva, que impide llegar a diagnosticar la enfermedad en aquellos individuos que mueren por otras causas. Por ello, los estudios de cohortes deben proporcionar información sobre el porcentaje y el tipo de pérdidas observadas a lo largo del seguimiento.

13.1 Las técnicas de análisis de supervivencia

Utilizan la información individual de cada participante de la cohorte. Las técnicas más conocidas son las curvas de supervivencia de Kaplan-Meier, cuya significación estadística es comprobada a través del test de logrank, y los modelos semiparamétricos de Cox. La popularidad de ambos métodos se basa en la ausencia de asunciones sobre la evolución de la tasa de incidencia en el tiempo. Sin embargo, en ambos casos (logrank y Cox) se asume que la medida del efecto, es decir, la razón de tasas de incidencia (denominada "Hazard Ratio" en los estudios de supervivencia) es constante a lo largo del tiempo (asunción de proporcionalidad). De todas formas, el modelo de Cox permite parametrizar la falta de proporcionalidad dentro del propio modelo. También es posible incluir dentro del modelo exposiciones cambiantes en el tiempo. En la figura 6-5 se muestra, como ejemplo, la comparación entre las curvas de supervivencia obtenidas en la cohorte de CPNM en enfermos con niveles preoperatorios bajos de antígeno carcinoembrionario (CEA) y en pacientes con niveles elevados. La diferencia entre ambas es estadísticamente significativa.

Para el análisis de supervivencia es importante que la unidad temporal esté referida al tiempo biológicamente relevante, que puede o no coincidir con la cronología del estudio. Por ejemplo, en el estudio MACS no tiene mucho sentido considerar el inicio del estudio como origen en el análisis de supervivencia, ya que los individuos seroprevalentes se encuentran en una fase diferente de la enfermedad, mientras que en los seronegativos el momento de la serconversión es el momento relevante.

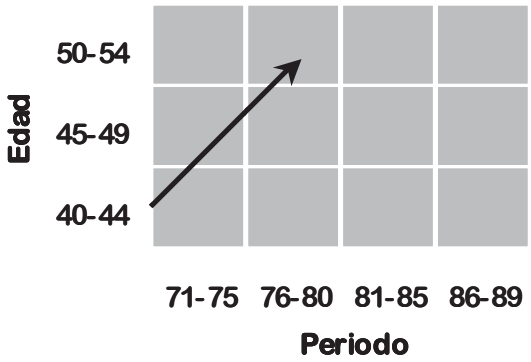
Figura 6-5: Comparación de las curvas de supervivencia postoperatoria entre los pacientes del estudio de factores pronósticos del cáncer de pulmón no microcítico según niveles séricos preoperatorios de antígeno carcinoembrionario.



13.2 Las técnicas de personas-año

Utilizan como unidad de análisis estratos o grupos homogéneos en los que es razonable asumir una misma tasa de incidencia. Esto requiere colapsar la información individual por estratos de interés. Estas técnicas son utilizadas en los grandes estudios poblacionales, como la cohorte de población activa sueca. Las variables que definen los estratos de interés son, lógicamente, la exposición y el resto de factores de riesgo (factores de confusión) a considerar. Algunas de estas variables, como el sexo, pueden ser fijas, mientras que otras (la edad o el periodo de observación, por ejemplo) cambian a lo largo del tiempo. La existencia de estas variables cambiantes en el tiempo hacen necesario "dividir" el tiempo de observación de cada persona para considerar su contribución a cada uno de los estratos (grupos de edad, por ejemplo) por los que va pasando. Si tomamos un sujeto de la cohorte sueca con 43 años en 1971 (ver figura 6-6) y que es diagnosticado de cáncer de mama en 1978, podemos ver, por ejemplo, que su contribución en personas-año correspondiente al periodo 71-75 debe ser dividida en dos porciones: una parte corresponde al grupo de edad de 40-44 años y otra parte al grupo de edad 45-49. En la cohorte sueca este proceso fue realizado en cada sujeto de la misma.

Figura 6-6: Contribución de un individuo a diferentes estratos en el cálculo de las personas año.



En cohortes históricas poblacionales, generalmente los estratos a considerar vienen determinados por las variables para las que se dispone de información. En estos estudios, como ocurre en los estudios descriptivos de grandes colectivos de población, con frecuencia el denominador de las tasas no se calcula de forma exacta, contabilizando la contribución de cada individuo, sino considerando la población a mitad del periodo multiplicada por la duración del periodo.

Las técnicas clásicas de análisis "personas tiempo" incluyen la comparación de tasas, la estandarización directa y las razones estandarizadas de incidencia o mortalidad (denominadas en la literatura RIE o SIR, para la incidencia y RME o SMR para la mortalidad). RIE y RME comparan el número de casos observados con el número esperado teniendo en cuenta la frecuencia de enfermedad en la población utilizada como referencia. La técnica multivariante por excelencia para la modelización de las tasas es la regresión de Poisson, que utiliza la distribución de Poisson para modelizar la variabilidad aleatoria del numerador de las tasas.

14. POTENCIA ESTADÍSTICA

El alto coste de los estudios de cohortes en términos de esfuerzo, tiempo y dinero hace necesario la estimación previa del tamaño muestral requerido en relación al efecto que se pretende observar. En muchas situaciones, sin embargo, es más práctico proceder al revés. Es decir, considerar en primer lugar el número de participantes que es posible incluir en el estudio, bien sea por restricciones económicas o de otro tipo, y calcular la potencia estadística disponible.

Por ejemplo, en el estudio sobre factores pronósticos de CPNM, el volumen de pacientes atendidos en los hospitales participantes indicaba la posibilidad de incluir alrededor de 400 casos durante el reclutamiento. Teniendo en cuenta la proporción de muestras p53 positivas en series previas de pacientes, dicho tamaño muestral suponía, con un periodo de seguimiento mínimo de dos años, la posibilidad de detectar significación estadística (potencia superior al 80%) para riesgos relativos (hazard ratios) iguales o superiores a 1,50.

Una vez concluido el estudio, el cálculo de la potencia estadística a posteriori es particularmente importante para enjuiciar los resultados negativos, es decir, la ausencia de asociación entre la exposición y la enfermedad. Los métodos estadísticos no permiten nunca demostrar la hipótesis nula (falta de asociación) sino sólo rechazarla o no con un margen de error determinado. Por ello, cuando la incidencia del evento de interés es similar en expuestos y no expuestos, es importante tener en cuenta las siguientes consideraciones:

— ¿Cuál es la potencia estadística real del estudio? ¿Cuál es el mínimo RR que sería posible detectar teniendo en cuenta las características del mismo?
— ¿Existen diferencias reales de exposición en los grupos de comparación?
— ¿El periodo de seguimiento ha sido suficiente? Hay que tener en cuenta el periodo de latencia necesario (ver figura 6-1).
— ¿Cómo encajan los resultados del estudio en el marco general del conocimiento existente sobre dicha enfermedad?

15. PROBLEMAS DE INTERPRETACIÓN DE LOS ESTUDIOS DE COHORTES

15.1 Representatividad de la cohorte seleccionada

Este problema puede estar presente cuando se seleccionan grupos específicos, por su disponibilidad o su mayor tasa de participación, para representar a toda la población (el estudio de los médicos británicos, o de las enfermeras americanas, por ejemplo). La inclusión en los estudios de voluntarios, es decir, personas que aceptan participar, también puede conllevar problemas de representatividad (además de los estudios citados anteriormente, el EPIC y el MACS utilizan voluntarios). Los voluntarios suelen ser personas con mayor nivel de educación, más preocupados por su salud y con hábitos de vida más saludables. Esto puede implicar una infraestimación de exposiciones nocivas y una menor incidencia de la enfermedad.

15.2 Elección del grupo de comparación

Las personas expuestas y no expuestas pueden diferenciarse respecto a otros factores relevantes para la enfermedad de interés. Cuando estos factores son conocidos, el estudio debe recoger información sobre ellos y obtener estimadores del efecto de la exposición, eliminando el posible efecto de confusión de estas variables. En estudios que utilizan grupos externos de comparación, esta información no suele estar disponible.

15.3 Pérdidas durante el seguimiento

Como ya se ha comentado, las técnicas de análisis consideran todos los individuos que no llegan a desarrollar la enfermedad como un grupo homogéneo, pero las causas de esta falta de información son diversas. El principal problema radica en los sujetos que abandonan el estudio y en aquellos que mueren por otras causas. Las técnicas analíticas asumen que las pérdidas son aleatorias y no están relacionadas con la enfermedad o la exposición de interés. Por ello, es importante comprobar que las personas que permanecen en el estudio y aquellas que lo han

abandonado son realmente similares. Cuando las pérdidas no son aleatorias se produce un sesgo de selección a lo largo del tiempo y la cohorte deja de representar a la población de base inicial.

15.4 Sesgos introducidos por los errores de medida

Ya se ha hablado de ellos en el apartado dedicado a la exposición (ver también la tabla 6-3). Es importante que dichos errores sean de tipo "no diferencial", es decir, que el error esperado sea similar en los distintos grupos de exposición.

15.5 El problema de las comparaciones múltiples

Este problema se presenta en aquellos estudios de cohortes que valoran múltiples exposiciones o múltiples eventos de interés. Por ejemplo, en la cohorte de población activa sueca, el interés se centra en conocer la incidencia relativa de un determinado tipo de cáncer en cada grupo ocupacional. Esto supone la realización de múltiples contrastes, comparando cada una de las 270 ocupaciones con el resto de la población activa. En estos estudios, la posibilidad de encontrar, por azar, asociaciones estadísticamente significativas aumenta considerablemente. Recientemente se han propuesto nuevos procedimientos para recalcular la significación estadística teniendo en cuenta el número de comparaciones realizadas, minimizando la "tasa de hallazgos falsos" (False Discovery Rate).

16. CASOS Y CONTROLES BASADOS EN ESTUDIOS DE COHORTES

Se trata de diseños de estudios caso-control que pretenden maximizar la información obtenida a partir de los estudios de cohortes minimizando el coste. Los principales diseños son dos: los casos y controles anidados a una cohorte y los estudios caso-cohorte. Estos diseños son especialmente importantes cuando el estudio incluye determinaciones analíticas costosas o muestras difíciles de obtener, como por ejemplo, biomarcadores de exposición o susceptibilidad.

16.1 Estudios de casos y controles anidados

Son estudios caso-control que incluyen todos los casos observados en la cohorte y uno o más controles emparejados con los casos por tiempo de seguimiento. Este apareamiento implica considerar la información disponible para los controles sólo hasta la fecha de diagnóstico de su caso correspondiente. Además, un control puede llegar a convertirse en caso, a medida que avanza el seguimiento. El análisis de los datos debe respetar el apareamiento y comparar cada caso con su grupo de controles.

16.2 Estudios caso-cohorte

Bajo este diseño se seleccionan como controles, de forma aleatoria, una fracción de todos los componentes de la cohorte original. Este tipo de diseño supone la consideración de la información proporcionada por la subcohorte seleccionada más la de todos los casos ocurridos fuera de dicho grupo. En este tipo de diseño, los métodos de análisis deben tener en cuenta el carácter diferencial de los casos incluidos y no incluidos en la subcohorte seleccionada.

17. ESTUDIOS DE COHORTES Y LOS ENSAYOS CLÍNICOS

La diferencia entre estos dos diseños es la asignación aleatoria de la exposición, propia de los ensayos clínicos. Dicha asignación aleatoria, o randomización, garantiza la comparabilidad de los grupos de exposición siempre que el tamaño muestral sea adecuado.

Por razones éticas obvias, los ensayos clínicos sólo se utilizan para valorar exposiciones o intervenciones potencialmente beneficiosas. En estas circunstancias, los ensayos clínicos son preferibles a los estudios de cohortes. La randomización sitúa a los ensayos clínicos en el marco de investigación propio de los estudios experimentales, donde el investigador puede controlar las condiciones en las que se produce el experimento.

Sin embargo, cuando se investiga la utilidad de nuevos tratamientos, los criterios de inclusión utilizados en los ensayos clínicos suelen implicar una gran selección de los pacientes. Con frecuencia, los grupos más vulnerables (pacientes con otras patologías subyacentes, edades avanzadas, grupos sociales desfavorecidos, etc) no forman parte de estos estudios. Además, las condiciones de aplicación del tratamiento, el seguimiento y el manejo de los casos dentro de un ensayo clínico pueden no ser representativos del contexto clínico general en el que se diagnostica y trata a los pacientes. Estas diferencias explican los resultados paradójicos encontrados, por ejemplo, cuando se evaluó la utilidad de los primeros agentes antirretrovirales. Mientras los ensayos clínicos mostraban un efecto beneficioso, la mortalidad de los casos con SIDA de estudios de cohortes como el MACS no mostraban diferencias sustanciales entre los pacientes tratados y no tratados. Hubo que esperar a la llegada de la nueva terapia potente antiriretroviral para que el impacto de dichos tratamientos fuese realmente visible en las cohortes de pacientes con SIDA.

Teniendo en cuenta lo anterior, es interesante matizar que los ensayos clínicos determinan la eficacia teórica de un determinado tratamiento. Como contrapartida, los estudios de cohortes representativas de pacientes juegan un papel importante en el estudio de la efectividad, es decir, el efecto real de dicho tratamiento aplicado en las condiciones normales de la práctica clínica.

En otro tipo de contexto, la aleatorización resulta inviable al existir grupos naturalmente expuestos y grupos no expuestos en la población. Los estudios de cohortes son, en este caso, la mejor aproximación de que dispone el epidemiólogo para acercarse a la calidad de la información generada por los ensayos clínicos.

18. CONCLUSIONES

Como se ha comentado en este capítulo, los estudios de cohortes suponen una herramienta esencial en la investigación epidemiológica. En ellos, la relación temporal causa-efecto está asegurada, siendo posible además investigar e incorporar en el análisis los cambios que se producen en la cohorte a lo largo del tiempo.

El diseño y desarrollo de un estudio de cohortes es generalmente un reto que conlleva muchas dificultades. Es necesario valorar la exposición y los cambios que se producen a lo largo del tiempo, retener a los participantes en la cohorte y mantener un sistema de seguimiento que permita determinar la incidencia de la enfermedad o enfermedades estudiadas.

Los estudios de cohortes han demostrado ser útiles no sólo a la hora de investigar los factores etiológicos más importantes de muchas enfermedades, sino también para caracterizar la evolución de epidemias como el SIDA, predecir el riesgo de enfermar a nivel individual (por ejemplo, las fórmulas que valoran el riesgo de enfermedad cardiovascular proceden de la cohorte Framingham y de estudios similares), investigar el efecto sinérgico de dos o más exposiciones, determinar el periodo de latencia, investigar la validez de los marcadores precoces de enfermedad, etc. Finalmente los estudios de cohortes complementan a los ensayos clínicos a la hora de valorar la utilidad de los nuevos agentes terapéuticos. Los estudios de cohortes han servido además de estímulo en el planteamiento de nuevas hipótesis y el desarrollo de nuevos métodos de análisis. La riqueza de la información aportada por estos estudios, la capacidad que han

demostrado para incorporar nuevas hipótesis y ayudar a clarificar el papel desempeñado por diferentes factores de riesgo en la génesis de enfermedades complejas sitúan a estos estudios como marco de referencia en la investigación epidemiológica.

19. BIBLIOGRAFÍA

BENJAMINI Y, HOCHBERG Y. *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J Royal Stat Soc B 1995; 57:289–300.

BRESLOW NE, DAY NE. *Statistical Methods in Cancer Research Volume II : The design and analysis of cohort studies.* IARC Scientific Publications No. 82. International Agency for Research on Cancer. Lyon, 1987.

COLDITZ GA, MANSON JE, HANKINSON SE: *The Nurses' Health Study: 20-year contribution to the understanding of health among women.* J Women's Health 1997; 6: 49-62.

DOLL R, HILL AB. *Mortality in relation to smoking: ten years' observations of British doctors.* Br Med J 1964; 1: 1399-1410.

GUALLAR EM. *Alternativas al diseño tradicional de los estudios de cohortes: estudios de casos y controles híbridos y estudios de cohortes y casos.* Rev Salud Pública 1991; 2: 151-165.

HUNTER, D.J. *Methodological issues in the use of biological markers in cancer epidemiology: cohort studies. En Toniolo P. y cols..* Ed. Application of Biomarkers in Cancer Epidemiology. IARC Scientific Publications Nº142. International Agency for Research on Cancer. Lyon, 1997.

KAUPPINEN T, TOIKKANEN J Y PUKKALA E. *From cross-tabulations to multipurpose exposure information systems: a new job-exposure matrix.* Am J Ind Med 1998; 33:409-417

LYLES RH, MUNOZ A, YAMASHITA TE, BAZMI H, DETELS R, RINALDO CR, MARGOLICK JB, PHAIR JP, MELLORS JW. *Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study.* J Infect Dis. 2000;181:872-80.

MUÑOZ A, HOOVER DR. *Use of cohort studies for evaluating AIDS therapies.* En: Finkelstein DM, Schoenfeld DA (eds). AIDS Clinical Trials. Wily and Sons. NY 1995. pp: 423-446.

NIETO-GARCÍA FJ. *Los Estudios de cohorte. En: Martínez-Navarro F, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V (eds).* Salud Pública. McGraw-Hill y Interamericana. Madrid, 1999.

Organización Internacional del Trabajo. Enciclopedia de Salud y Seguridad en el Trabajo. 3ª Ed. Ministerio de Trabajo y Asuntos Sociales 2001. http://www.mtas.es/insht/EncOIT/Index.htm.

POLLÁN M, GUSTAVSSON P. *Cancer and Occupation in Sweden: 1971-1989.* EPC-Rapport 1999:1.The National Board of Health and Welfare, Sweden, 1999. Disponible en: http://www.sos.se/fulltext/9918-001.pdf.

POLLÁN M. *Ocupación, exposición laboral a radiaciones electromagnéticas y cáncer de mama.* Instituto de Salud Carlos III. Madrid, 2001. Disponible en: http://193.146.50.130/cancer/cem-cancer.pdf.

POLLÁN M, VARELA G, TORRES A, DE LA TORRE M, LUDEÑA MD, ORTEGA MD, PAC J, FREIXENET J, GÓMEZ G, SEBASTIÁN F, DÍEZ M, ARRABAL R, CANALÍS E, GARCÍA-TIRADO J, ARNEDILLO A, RIVAS JJ, MINGUELLA J, GÓMEZ A, GARCÍA M, ARAGONÉS N, PÉREZ-GÓMEZ B, LÓPEZ-ABENTE G, GONZÁLEZ-SARMIENTO R, ROJAS JM. *Clinical value of p53, c-erbB-2, CEA and CA125 regarding relapse, metastasis and death in resectable non-small cell lung cancer.* Int J Cancer 2003; 107: 781-90.

RIBOLI E, KAAKS R. The EPIC *Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol 1997 26: S6-S14.

ROTHMAN KJ, GREENLAND S. *Cohort Studies.* En: Modern Epidemiology. Second Edition. Lippincott-Raven. Philadelphia, 1998. pp 79-91.

SAMET JM, MUÑOZ A (EDS). *Cohort Studies.* Epidemiologic Reviews 1998; 20, 1.

SCHOTTENFELD D, FRAUMENI FJ, JR. *Cancer Epidemiology and Prevention.* Second Edition. Oxford University Press. Oxford, 1996.

TARWATER PM, MELLORS J, GORE ME, MARGOLICK JB, PHAIR J, DETELS R, MUNOZ A. *Methods to assess population effectiveness of therapies in human immunodeficiency virus incident and prevalent cohorts.* Am J Epidemiol. 2001;154: 675-81.

VII. ESTUDIOS DE CASOS Y CONTROLES

F. Javier Jiménez Jiménez

1. INTRODUCCIÓN

Un estudio de casos y controles es un estudio analítico observacional que comienza con la identificación de un grupo de casos (individuos con una particular enfermedad o condición) y un grupo de controles (individuos sin la enfermedad o condición). Entonces, el nivel o prevalencia de la exposición a un factor se mide en los dos grupos y se compara. Básicamente, si la prevalencia de la exposición es mayor en casos que en controles, la exposición podría ser un factor de riesgo, mientras que si es menor, la exposición podría ser un factor protector.

Aunque tradicionalmente los estudios de casos y controles se conocían también como estudios retrospectivos, debido a que el muestreo se realiza a partir de la situación caso/control de los participantes, como se ha señalado en un capítulo precedente estos estudios pueden ser tanto prospectivos como retrospectivos. Esta terminología, sin embargo, puede dar lugar a confusión con los estudios de cohorte retrospectivos, por lo que su uso tiende a abandonarse.

El papel de los estudios de casos y controles en la evaluación de hipótesis epidemiológicas ha estado limitado por dificultades en cuanto a la interpretación del tipo de evidencia aportada. Frecuentemente se considera que los estudios de casos y controles recogen la evidencia de forma contraria a lo que sucede en la vida real y que son, en cierto modo, 'anti-naturales'. Esto es así porque, al contrario que los estudios de cohortes, abordan el proceso de enfermar partiendo del estado final (enfermedad) para llegar a la fase inicial (exposición). Actualmente está más extendida la idea de que los estudios de casos y controles son, en realidad, una forma eficiente de muestrear una cohorte subyacente.

Para entender los estudios de casos y controles, la aproximación más útil es considerar que existe una población de base, a partir de la cual obtenemos los casos y en la que muestreamos entre las personas a riesgo para obtener los controles. Así, los casos aportan información con respecto a la exposición en las personas que desarrollan la enfermedad, mientras que los controles aportan información acerca de la exposición en la población de la que proceden los casos.

Aunque los estudios de casos y controles están sujetos a algunas fuentes de sesgos diferentes de los estudios de cohorte (por ejemplo, el hecho de que la enfermedad modifique la exposición es un problema fundamentalmente en los estudios de casos y controles), desde el punto de vista lógico, no se diferencian de los estudios de cohorte.

2. USOS DE LOS ESTUDIOS DE CASOS Y CONTROLES. VENTAJAS Y DESVENTAJAS

Los estudios de casos y controles se han utilizado principalmente en la investigación de los factores asociados con el desarrollo de enfermedades crónicas (cáncer, enfermedades cardiovasculares), aunque cada vez son más utilizados también en el estudio de enfermedades transmisibles y en la evaluación de intervenciones y programas de salud.

A continuación se detallan las principales ventajas y desventajas de los estudios de casos y controles, sobre todo en comparación a los estudios de cohorte.

Ventajas de los estudios de casos y controles

Los estudios de casos y controles son más baratos y rápidos que los estudios de cohorte. En los estudios de casos y controles se obtienen todos los casos que aparecen en una población de base teórica en el momento de sufrir la enfermedad. En este sentido, no es necesario realizar el seguimiento de esta población, por lo que los estudios de casos y controles son más sencillos desde el punto de vista logístico y no están sujetos a problemas de pérdidas en el seguimiento.

Los estudios de casos y controles son especialmente útiles en enfermedades raras o con largos periodos de latencia. En este caso, los estudios de cohorte son ineficientes, ya que es necesario seguir a un número elevado de personas durante bastante tiempo para obtener un número suficiente de casos. En muchas ocasiones, los estudios de casos y controles son la única opción realista. De hecho, existen numerosas enfermedades con incidencias inferiores a 1 caso / 10.000 personas-año. En esos casos, si deseáramos investigar las causas de una de esas enfermedades mediante un estudio de cohortes, habría que seguir a 100.000 personas durante 2 años para obtener menos de 20 casos de enfermedad por término medio, lo que resulta muy ineficiente.

Los estudios de casos y controles permiten estudiar una amplia variedad de posibles exposiciones. Aunque esto es también posible en estudios de cohorte, en éstos habría que determinar la exposición en todos los participantes de la cohorte, lo que puede ser ineficiente o, sencillamente muy caro. De hecho, estas consideraciones sobre la eficiencia en los estudios de cohorte llevaron al desarrollo de los diseños de casos y controles anidados y a otros diseños híbridos de muestreo dentro de una cohorte.

Desventajas de los estudios de casos y controles

Los estudios de casos y controles no son adecuados cuando el desarrollo de la enfermedad altera los niveles de las exposiciones que se están intentando determinar. En ocasiones, puede ser difícil determinar si la exposición ha causado la enfermedad, o si la enfermedad ha modificado la exposición. Por ejemplo, durante la fase aguda de un infarto de miocardio, los niveles de colesterol total en sangre disminuyen. Por lo tanto, si se realiza un estudio de casos y controles comparando los niveles de colesterol en casos incidentes de infarto agudo de miocardio con controles sin la enfermedad, los niveles de colesterol total serán más bajos en los casos, a pesar de que numerosos estudios prospectivos han establecido que los sujetos con niveles más elevados de colesterol total tienen un mayor riesgo de padecer un infarto de miocardio.

Los estudios de casos y controles son más propensos a sesgos de selección y de información. Por un lado, la selección de la muestra puede afectar la asociación resultante. De hecho, una de las dificultades principales en los estudios de casos y controles es la selección de un grupo control adecuado (ver selección de controles más adelante). Por otro lado, la información sobre la exposición se recoge después de que los casos hayan desarrollado la enfermedad, por lo que frecuentemente los sujetos de estudio y los investigadores saben qué participantes son casos y quienes controles y pueden recoger o recordar la información de forma diferente.

Los estudios de casos y controles no son eficientes para estudiar exposiciones raras. En este caso, tan sólo un número reducido de casos y de controles estarán expuestos, por lo que las estimaciones serán imprecisas.

Los estudios de casos y controles permiten estudiar tan sólo una enfermedad (o muy pocas) a la vez. En los estudios de cohorte, una vez reclutada la cohorte y seguida en el tiempo, es posible registrar la ocurrencia de múltiples enfermedades, por lo que se puede investigar la asociación entre las exposiciones determinadas en el estudio y numerosas condiciones. Los estudios de casos y controles, por el contrario, disponen habitualmente de tan sólo una serie de

casos, por lo que será necesario hacer un estudio de casos y controles para cada enfermedad que queramos investigar.

En los estudios de casos y controles no es posible, en general, obtener estimadores de la incidencia de la enfermedad. Tan sólo proporcionan medidas relativas de efecto (en concreto, odds ratios).

En los estudios de casos y controles no es posible actualizar la medida de la exposición (tomar medidas repetidas). En los estudios de cohortes es frecuente realizar varias visitas a los participantes durante el seguimiento, por lo que se puede volver a medir la exposición y utilizar esta información en el análisis del estudio.

3. PLANIFICACIÓN Y DISEÑO

Como todo estudio, un estudio de casos y controles requiere una planificación y organización cuidadosas. Será necesario definir claramente los objetivos y las hipótesis, detallar la metodología, establecer un cronograma realista y preparar un manual de operaciones para que el equipo de investigadores trabaje con criterios consistentes.

3.1. Definición de población de base

La población de base corresponde a la cohorte teórica de la que van a surgir tanto los casos como los controles. Debe ser concreta hasta donde sea razonable (edad, sexo, límites geográficos, etc.). Se han de establecer unos criterios de elegibilidad (inclusión/exclusión) que se deben aplicar tanto a casos como controles y que permiten restringir el estudio a las personas potencialmente en riesgo de la exposición. En el estudio EURAMIC sobre antioxidantes e infarto de miocardio, la población de base fueron los varones de menos de 70 años, nativos de los países del estudio, residentes en las zonas de captación, que hablaran la lengua oficial, y que no tuvieran historia previa de infarto de miocardio.

3.2. Definición y selección de casos

Al igual que en otros diseños epidemiológicos, es esencial tener criterios precisos para la definición de caso. Los criterios pueden estar basados en resultados de pruebas diagnósticas o de laboratorio (por ejemplo, ciertos datos histológicos para establecer el diagnóstico de un tipo de cáncer) o pueden estar basados en la historia clínica o en encuestas. Según la definición de los casos, se diferencian dos tipos importantes de estudios de casos y controles:

— Estudios de casos y controles con casos incidentes. Los casos del estudio se limitan a los casos nuevos de la enfermedad que aparecen durante el periodo de estudio. Siguiendo con el concepto de la equivalencia lógica entre los estudios de cohorte y los de casos y controles, esta situación reproduce lo que encontraríamos en un estudio de cohortes, y será, por tanto, el método de elección para la selección de casos en los estudios de casos y controles.

— Estudios de casos y controles con casos prevalentes. Los casos son todos aquellos pacientes que tienen la enfermedad en un momento determinado. El problema de los casos prevalentes es que la prevalencia depende tanto de la ocurrencia de la enfermedad (incidencia) como de su duración, que a su vez depende del tratamiento y de otros factores pronósticos. Por tanto, cuando se estudian series con casos prevalentes, es difícil saber si los factores que se asocian a la condición de casos han modificado la incidencia de la enfermedad o tan sólo su pronóstico. Además, los casos prevalentes no son representativos de los casos incidentes, ya que están sobre-representados los casos

de mayor duración. Desde el punto de vista del estudio de factores de riesgo (o protectores), los estudios de casos y controles con casos prevalentes son similares a los estudios transversales.

Además del tipo de casos (incidentes o prevalentes), es necesario considerar de dónde proceden los casos. Algunas alternativas son:

— Estudio de casos hospitalarios, en el que los casos se obtienen de los pacientes que cumplen los criterios de la definición de caso y que son atendidos en un determinado hospital. En este contexto habrá que tener en cuenta el conocido como "sesgo de Berkson", que se refiere a que los pacientes con varias enfermedades tienen una mayor probabilidad de ser hospitalizados que aquellos que sólo padecen una, por lo que los casos seleccionados tendrán una mayor proporción de pacientes con una segunda enfermedad que un grupo extraído de la población general.

— Estudio de casos poblacionales, en el que los casos se toman de una población definida, habitualmente a partir de un registro poblacional de la enfermedad en estudio.

En cualquier caso, es importante considerar si los casos elegidos son representativos de todos los casos en la población de base. Existe siempre un cierto grado de selección, y por tanto el papel del sesgo de selección en un estudio de casos y controles es particularmente importante. Debe asegurarse que la selección de casos (e igualmente de controles) se realice independientemente de su estatus de exposición. En el caso concreto de la selección de casos es importante asegurarse que la situación de expuesto no determina una búsqueda más exhaustiva de la enfermedad en estudio.

3.3. Definición y selección de controles

Los controles deben cumplir los criterios de definición de casos excepto los relativos a la enfermedad. En general un control debe ser una persona que si hubiera desarrollado la enfermedad hubiera sido seleccionado como caso. Por ejemplo, si los casos son hospitalarios, un control será un sujeto que de haber desarrollado la enfermedad hubiese ido a parar al hospital del que proceden los casos.

La población de base de los controles debe ser la misma que la de casos y tanto los casos como los controles deben ser seleccionados concurrentemente. Además, es necesario tener en cuenta que el tiempo durante el cual un individuo es susceptible de ser seleccionado como control debe ser el tiempo durante el que ese individuo es susceptible de ser seleccionado como caso. De esta forma, un individuo que desarrolla la enfermedad en estudio o que muere no puede ser ya seleccionado como control. Igualmente, la probabilidad de seleccionar cualquier potencial control debe ser proporcional al tiempo que dicho control aportaría al denominador del calculo de una tasa si se realizara un estudio de cohortes.

Según su procedencia, los controles pueden ser:

— Poblacionales. En este caso, los controles son una muestra aleatoria directamente obtenida de la población de base. En teoría, los controles poblacionales proporcionan la mejor representación de la exposición en la población de la que proceden los casos, aunque la menor participación de los mismos puede limitar la credibilidad del estudio.

— Hospitalarios. En ocasiones, la población de base no es fácilmente identificable, de forma que una muestra aleatoria de la población general no representa necesariamente una muestra de la población fuente de los casos (especialmente, si los casos son hospitalarios). Además, en muestreos poblacionales es frecuente obtener bajas tasas de respuesta, sobre todo si se realiza alguna medida invasiva. En estos casos es conveniente

la selección de controles a partir de los mismos hospitales de los que surgen los casos. El principal problema surge al intentar seleccionar los controles de manera independiente de la exposición (a menudo, los sujetos expuestos tienen diferente probabilidad que los no expuestos de acudir al hospital por diferentes enfermedades). Con objeto de evitar este sesgo se deben seleccionar como controles individuos afectos de enfermedades que, en principio, no están asociadas con la enfermedad, así como seleccionar un amplia variedad de enfermedades como control.

— Sujetos fallecidos. Si los casos utilizados en el estudio se definieran como muertes por cierta causa, o si algunos casos fallecieran entre el diagnóstico y el momento de ser entrevistado, se podría plantear la selección de controles fallecidos para buscar una mayor comparabilidad. Sin embargo, el uso de controles fallecidos puede llevar a una evaluación sesgada de la exposición en la población de la que surgen los casos si dicha exposición causa o previene la mortalidad o se asocia con otros factores que actúan sobre la mortalidad. La principal justificación para el uso de controles fallecidos es la conveniencia, como en el caso de estudios basados completamente en sujetos fallecidos (estudios de mortalidad proporcional).

— Vecinos. El muestreo de vecinos como controles puede ser una alternativa a la selección de controles poblacionales cuando no es posible realizar un muestreo probabilístico de la población. En este caso, es posible que los casos y controles, además de vecindario, compartan exposiciones similares, lo que podría resultar en sesgos.

— Sujetos nombrados por los casos (compañeros o amigos). Este tipo de selección determina la existencia de un emparejamiento individual. Además de las ventajas e inconvenientes inherentes al emparejamiento, la selección de amigos o vecinos como controles presenta como inconvenientes que la selección puede estar asociada con la exposición (por residencia cercana o por hábitos similares) y que la decisión de la selección de controles puede depender en exceso de los casos (amigos).

En la selección de controles hay que evitar los sesgos de selección, es decir, evitar que la selección de controles esté determinada por el estatus de exposición, ya que es fácil elegir inadvertidamente un grupo control con una prevalencia de exposición mayor o menor que la población general. Así, en los estudios de la asociación entre el alcohol y el cáncer de mama, el uso de controles a partir de pacientes que acuden a urgencias (del mismo hospital que los casos) puede producir un sesgo dado que este grupo tiene un mayor consumo de alcohol que la población general (los pacientes acuden a urgencias por accidentes relacionados frecuentemente con el consumo de alcohol).

3.4. Medida de la exposición en casos y en controles

Los datos referentes a la exposición pueden ser obtenidos de diversas maneras, destacando la entrevista (personal, postal o telefónica), entrevistas con parientes, examen de registros médicos, ocupacionales o de otro tipo, o la realización de exploraciones o toma de muestras biológicas.

Independientemente del método de recogida de la información, es necesario tener en cuenta dos aspectos. Por un lado, la información recogida no debe estar influenciada por el hecho de que un individuo sea caso o control. Por otro lado, es importante asegurar la validez del método utilizado para evaluar la exposición.

La falta de comparabilidad en la información de casos y controles pueda dar lugar a la presencia de sesgos de información, entre los que destacan:

— Sesgo del observador. Se produce cuando los datos correspondientes a la exposición son recogidos de manera diferente para los casos que para los controles. Los casos

pueden ser entrevistados de manera más exhaustiva sobre la exposición que los controles. Es conveniente que el investigador no conozca las hipótesis del estudio o, mejor todavía, si el paciente es un caso o un control. La información referente a la exposición debe recogerse de igual manera en casos y controles y usando los mismos instrumentos en ambos.

— Sesgo de respuesta. Este sesgo se refiere al proceso por el cual la información aportada con respecto a la exposición es diferente en casos y controles. Un ejemplo es el sesgo de memoria, tener una determinada enfermedad puede hacer que se recuerden mejor las posibles exposiciones. Este sesgo se intenta controlar evitando que los participantes en el estudio conozcan las hipótesis del mismo e intentando conseguir que casos y controles tengan el mismo incentivo para recordar exposiciones pasadas.

La falta de validez en los métodos de medida de la exposición, no asociada al status caso/control, da lugar a un error de clasificación no diferencial que, en general, tenderá a sesgar los resultados del estudio hacia el valor nulo, es decir hacia la falta de asociación entre la exposición y la enfermedad o condición evaluada.

4. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

4.1. Odds ratio y riesgo relativo

El análisis de los estudios de casos y controles se basa en los conceptos de odds y odds ratios y se realiza mediante la comparación de casos y controles con respecto a la frecuencia de exposición a través del odds ratio de la exposición, que es el odds de la exposición entre los casos dividido por el odds de la exposición entre los controles.

Sin embargo, para evaluar la relación de una determinada exposición con el desarrollo de una enfermedad es más fácil entender la asociación a partir de la posibilidad de desarrollar la enfermedad según se esté expuesto o no, que a partir de la probabilidad de estar expuesto según se tenga la enfermedad o no. Así, si seguimos dos poblaciones (expuesta y no expuesta) durante un determinado tiempo (t), podemos calcular la asociación de la exposición con el desarrollo de la enfermedad mediante el riesgo relativo, como la razón entre la tasa de incidencia en expuestos (probabilidad de desarrollar la enfemedad en el tiempo t entre los expuestos) y la tasa de incidencia en no expuestos.

Tal y como se ha comentado anteriormente, los estudios de casos y controles pueden ser vistos como un esquema de muestreo eficicente de la "experiencia de la enfermedad" de una cohorte subyacente, lo que permite una visión integrada de los estudios de casos y controles y de cohortes.

Consideremos, por ejemplo, una población dinámica que se encuentra en un una situación estable durante una cierta fracción de tiempo en la que es observada. En este tiempo, x_e individuos expuestos y x_{ne} individuos no expuestos serán diagnosticados por primera vez de una determinada enfermedad. Si conocemos el tiempo que cada individuo (expuesto o no expuesto) que se encontraba sano al inicio del periodo de observación ha permanecido en riesgo de desarrollar la enfermedad, podremos calcular la tasa de incidencia de la enfermedad entre los expuestos y la tasa de incidencia de la enfermedad entre los no expuestos, y por tanto el riesgo relativo. Sin embargo, esta información no se puede obtener en los estudios de casos y controles. Para calcular la densidad o tasa de incidencia, es necesario conocer el número de personas-tiempo a riesgo tanto para expuestos como para no expuestos. Si consideramos τ_e el total de personas

tiempo aportado por los individuos expuestos y τ_{ne} el total de personas tiempo aportado por los individuos no expuestos, el riesgo relativo de la enfermedad se calcularía como $(x_e/\tau_e)/(x_{ne}/\tau_{ne})$, que puede ser escrito igualmente como $(x_e/x_{ne})/(\tau_e/\tau_{ne})$.

En un estudio de casos y controles basado en la misma población, la selección adecuada de los casos, de manera que sean representativos del total de casos ocurridos en ese tiempo, permite determinar directamente el ratio x_e/x_{en}. Por otro lado, la razón τ_e/τ_{ne} se puede estimar de manera no sesgada y sin necesidad de asunciones sobre la frecuencia de la enfermedad a partir de la razón del número medio de individuos a riesgo expuestos *versus* no expuestos en la población de base. Tanto τ_e como τ_{ne} se pueden estimar como el producto del periodo de observación por el número medio de personas a riesgo. Dado que el periodo de observación es el mismo para expuestos que para no expuestos, la razón τ_e/τ_{en} es equivalente a la razón entre el número medio de personas expuestas a riesgo y el número medio de personas no expuestas a riesgo. De esta forma, en un estudio de casos y controles, la selección de un grupo control que refleje adecuadamente el ratio entre el número medio de personas expuestas a riesgo y el número medio de personas no expuestas a riesgo permitiría que el *odds* de la exposición en los controles pudiera verse como una estimación no sesgada de la razón de personas-tiempo expuestas *versus* no expuestas, y el estudio de casos y controles como un proceso de muestreo eficiente para estudiar el riesgo relativo de la enfermedad en la cohorte subyacente.

4.2. Estrategia para el análisis de un estudio de casos y controles

El análisis estadístico de un estudio de casos y controles se corresponde al análisis de datos binarios (dicotómicos): la variable dependiente (caso o control) se estudia en función de una serie de variables independientes (exposiciones de interés, factores de confusión y modificadores de efecto). De forma sistemática, el análisis de un estudio de casos y controles y la presentación de sus resultados puede desarrollarse según el siguiente esquema:

— Comparación de las variables demográficas, exposiciones y otras variables entre casos y controles, utilizando procedimientos estadísticos simples (t de Student, χ^2).

— Estudio de la asociación entre la variable de exposición principal y las otras variables en el grupo control (que representa a la población de la que proceden los casos), utilizando técnicas de regresión lineal, correlación o tablas de contingencia). Este análisis permite identificar potenciales factores de confusión que deberán ser tenidos en cuenta en análisis posteriores.

— Análisis estratificado de la asociación entre el desarrollo de la enfermedad y la variable de exposición principal (análisis de tablas de contingencia).

— Análisis multivariante de la asociación entre el desarrollo de la enfermedad y la variable de exposición principal (regresión logística en estudios de casos y controles no emparejedos y regresión logística condicional en estudios de casos y controles emparejados).

En resumen, los estudios de casos y controles permiten abordar prácticamente cualquier investigación abordable por un estudio de cohortes (salvo aquellas situaciones en las que la enfermedad en estudio puede influenciar la exposición evaluada. No obstante, los estudios de casos y controles requieren un cuidadoso diseño para evitar los sesgos de selección e información a los que es más propenso que los estudios de cohortes.

5. BIBLIOGRAFÍA

BRESLOW NE, DAY NE. *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies.* Lyon: International Agency for Research on Cancer, 1980.

CORREA A, STEWART WF, YEH H-C, SANTOS-BURGOA C. *Exposure measurement in case-control studies: reported methods and recommendations.* Epidemiol Rev. 1994;16:18-32.

LASKY T, STOLLEY PD. *Selection of cases and controls.* Epidemiol Rev 1994;16:6-17.

MACMAHON B, TRICHOPOULOS D. *Epidemiology. Principles & Methods. 2ª ed.* Boston: Little, Brown and Company, 1996.

ROTHMAN KJ, GREENLAND S. *Modern Epidemiology. 2ª ed.* Filadelfia: Lippincott-Raven; 1998.

SCHLESSELMAN JJ. *Case-Control Studies. Design, Conduct, Analysis.* Nueva York: Oxford University Press, 1982.

SZKLO M, NIETO FJ. *Epidemiology. Beyond the basics.* Gaithersburg: ASPEN, 2000

THOMPSON WD. *Statistical analysis of case-control studies.* Epidemiol Rev 1994;16:33-50.

VIII. ESTUDIOS DE PREVALENCIA

Juan de Mata Donado Campos

1. INTRODUCCIÓN

El estudio de las enfermedades en las poblaciones humanas se lleva a cabo de dos maneras.

1. Describiendo la forma en que la enfermedad se distribuye en la comunidad. Para ello intentamos responder a las siguientes preguntas:

- ¿Cuándo aparece la enfermedad? Estudio del Tiempo
- ¿Por dónde se distribuye? Estudio del lugar
- ¿A quién afecta? Estudio de las personas.

2. Analizando las causas o los factores de riesgos que hacen que aparezca la enfermedad. Es decir, intentamos responder a las preguntas

- ¿Por qué aparece la enfermedad en una determinada época del año o por qué aparece durante todo el año?
- ¿Por qué aparece en unas áreas y no en otras?
- ¿Por qué afecta sólo a unas determinadas edades o afecta a un determinado sexo?

La forma en que respondemos a estas preguntas es mediante la realización de estudios epidemiológicos.

1. La respuesta al "cuándo", al "dónde" y al "quién" la realizan los estudios descriptivos.

- Si estudiamos a poblaciones utilizamos los estudios ecológicos.
- Si estudiamos a personas utilizamos los estudios de prevalencia o transversales.

2. La respuesta a los "porqués" la realizan los estudios analíticos.

- Si podemos influir sobre las causas o los factores de riesgo (o de protección) de la enfermedad utilizamos los estudios de intervención comunitaria.
- Si no podemos influir sobre estos factores utilizamos los estudios observacionales: cohortes y de casos y controles. Bajo ciertas condiciones, que veremos a lo largo de este capítulo, también se incluyen a los estudios de prevalencia o transversales.

2. CARACTERÍSTICAS DEL DISEÑO

Los estudios de prevalencia son un tipo de investigación observacional descriptiva y analítica, bajo ciertos supuestos que veremos, en el que en un único momento temporal medimos la frecuencia de la exposición y del efecto (prevalencia). El tipo de diseño epidemiológico que se utiliza es el transversal, por lo que los términos transversal y prevalencia se suelen utilizar como sinónimos. Estos estudios son considerados por algunos autores, en los casos de exposiciones que varían con el tiempo, como estudio de casos y controles en los que se asume que el tiempo que media entre la exposición y el inicio de la enfermedad es nulo.

Las diferencias principales con los otros estudios observaciones son las siguientes.

1. Estudios de cohortes o de seguimiento
   - Existen dos poblaciones: expuestos y no expuestos
   - Existe un seguimiento: prospectivo o retrospectivo
   - Se miden incidencias y por tanto se pueden estimar riesgos. La medida del efecto es el riesgo relativo.
2. Estudios de casos y controles
   - Existen dos poblaciones: casos y controles
   - No existe seguimiento, aunque se pueden estudiar casos incidentes.
   - No se pueden estimar riesgos. La medida del efecto es la *odds ratio.*
3. Estudios de prevalencia o transversales
   - Seleccionamos una muestra de una población determinada. Las técnicas de muestreo mas utilizadas son el muestreo aleatorio simple o el muestreo por conglomerados. Se pueden utilizar cualquier otro método: listín telefónico u otras listas, casos consecutivos, etc. En cualquier caso debemos controlar que el hecho de presentar o no el fenómeno que queremos analizar no influya sobre la probabilidad de ser seleccionado.
   - En la muestra seleccionada desconocemos quiénes están expuestos o no o quiénes que están enfermos o sanos.
   - Sin realizar ningún seguimiento se interroga a los individuos sobre su estado de exposición y enfermedad y se clasifican en las distintas categorías: enfermos-expuestos, enfermos-no expuestos, sanos-expuestos, sanos-no expuestos.
   - Se estiman la frecuencia (prevalencia) de cada categoría. Las medidas de los efectos son dos: la razón de prevalencia (aproximación al riesgo relativo de los estudios de cohortes) o la *odds ratio* de prevalencia (aproximación a la *odds* ratio de los estudios de casos y controles)

Un ejemplo típico de un estudio de prevalencia o transversal es la Encuesta Nacional de Salud de España. A continuación resumimos la metodología utilizada en la Encuesta realizada en 2006.

La Encuesta Nacional de Salud de España 2006/2007 (ENSE-06) fue una investigación, de carácter bienal, sobre la salud en su sentido más amplio y sus factores determinantes desde la perspectiva de los ciudadanos. La ENSE-06 tuvo como objetivo general proporcionar la información necesaria sobre la salud de la población para poder planificar y evaluar las actuaciones en materia sanitaria. Los objetivos específicos fueron: 1. Proporcionar información sobre la valoración del estado de salud general, física y psíquica, e identificar los principales problemas que sienten los ciudadanos (enfermedades crónicas, dolencias, accidentes, limitaciones para realizar las actividades de la vida diaria) 2. Conocer el grado de acceso y utilización de los servicios de salud y sus características. 3. Conocer los factores determinantes de la salud: características del medio ambiente (físico y social) y hábitos de vida que suponen riesgo para la salud. 4. Realizar el análisis de la salud desde la perspectiva de género. 5. Conocer las desigualdades en salud.

La ENSE-06 aportó información esencial sobre trabajo reproductivo (sólo para personas adultas), estado de salud y enfermedades crónicas, accidentes, restricción de la actividad, consumo de medicamentos, salud mental, estrés laboral (sólo para personas adultas), acceso y

utilización de los servicios sanitarios, hábitos de vida, prácticas preventivas, agresiones, discriminación, características físicas y sensoriales, limitaciones para realizar las actividades de la vida diaria, apoyo afectivo y personal (sólo para personas adultas), y función familiar (sólo para personas adultas). También se investigaron las condiciones y entorno físico de la vivienda.

Además, para las personas menores de 16 años se obtuvo información sobre calidad de vida relacionada con la salud, tipo de lactancia (menores de 5 años) y tiempo dedicado a ver la televisión, Internet y vídeo-juegos. Con objeto de poder analizar los determinantes sociales de la salud, se obtuvo información sobre características sociodemográficas de la persona seleccionada, de su pareja (si convive con ella) y de la persona de referencia del hogar (persona que más aporta al presupuesto del hogar) La realización del estudio realizó en dos fases: 1. La primera se identifica con el Cuestionario de Hogar. Se intentó captar a todas las personas residentes en el hogar, solicitando a todos sus miembros información sobre algunas variables sociodemográficas fundamentales. Se seleccionó la persona adulta que debe contestar el cuestionario de salud de adultos y si en el hogar reside alguna persona menor de 16 años se seleccionó a de ellas para responder el cuestionario de salud de menores. Se solicitó para la persona adulta seleccionada y, en su caso, la persona menor seleccionada información sobre variables sociodemográficas adicionales y a su pareja (si convive con ella) y a la persona de referencia del hogar (si no coincide con alguna de las anteriores) otros datos también de carácter sociodemográfico. Por último se preguntó por características de la vivienda y del hogar.

2. La segunda fase se identificó por Cuestionario de Adultos (personas de 16 y más años) y el Cuestionario de Menores (personas de 0 a 15 años) En esta fase se recogió información de una única persona de 16 y más años, seleccionada aleatoriamente dentro del hogar, a través del Cuestionario de Adultos y de una única persona menor (personas de 0 a 15 años) de cada hogar si las hubiere, seleccionada aleatoriamente dentro del hogar a través del Cuestionario de Menores. La investigación se dirigía al conjunto de personas que residía en viviendas familiares principales. Cuando una misma vivienda estaba constituida por dos o más hogares, el estudio se extiende a todos ellos, pero de manera independiente para cada hogar. La Encuesta se realizó en todo el territorio nacional. El período de recogida de la información se extendió a lo largo de un año, desde junio de 2006 hasta junio de 2007.

El tipo de muestreo utilizado es un muestreo polietápico estratificado. El método de recogida de información fue el de entrevista personal la cuál podía ser complementada, en casos excepcionales, mediante entrevista telefónica. En cada vivienda se realizaron las visitas necesarias para obtener la información requerida. El trabajo de campo fue organizado a través de las 52 Delegaciones Provinciales del INE. Respecto al informante del cuestionario de adultos no se permitió información *proxy* salvo que la persona seleccionada no pudiese contestar por problemas de enfermedad, discapacidad o por estar internada en un centro sanitario y no vaya a regresar al hogar durante el tiempo que dure el trabajo de la sección.

Otro ejemplo de estudio de prevalencia es el proyecto EPINE. Este proyecto consiste en desarrollar una vez al año una encuesta transversal en los hospitales de enfermos agudos con objeto de determinar la prevalencia de infección nosocomial.

3. VENTAJAS

1. Como en un estudio de cohorte, podemos calcular la probabilidad de que una persona expuesta a un determinado factor de riesgo presente la enfermedad en un determinado momento.

2. Como en un estudio de casos y controles, podemos estimar la probabilidad de que una persona que está enferma tenga, o haya tenido, la exposición de nuestro interés.
3. Como en un estudio de cohortes, podemos estudiar varias enfermedades.
4. Se pueden realizar en un corto período de tiempo. Lo que implica un ahorro de tiempo y dinero. Dependiendo del tipo de muestreo realizado, se puede conseguir un buen control de la selección de los sujetos de estudio.
5. Suelen estudiar muestras representativas de la población. Esto favorece la extrapolación de los resultados a la población general como hemos visto en la Encuesta Nacional de Salud de España.
6. Son el primer paso en el diseño de un estudio de cohortes.
7. Proporcionan estimadores de prevalencia.

4. LIMITACIONES

1. Al estudiar sólo casos prevalentes, no tenemos seguridad de que sean representativos de la población general. Ya que los factores de riesgo asociados a los casos prevalentes pueden ser distintos a los asociados a los casos incidentes.
2. Podemos cometer un sesgo si utilizamos muestras procedentes de poblaciones no representativas de todos los casos. Por ejemplo, si deseamos estudiar la prevalencia de hepatitis B y seleccionamos una muestra de la población general y no incluimos a los usuarios de drogas ilegales podemos obtener una infraestimación de la prevalencia de dicha enfermedad.
3. Son ineficientes para estudiar enfermedades poco frecuentes, agudas o letales ya que la probabilidad de encontrar un caso en un momento dado (prevalencia) es muy baja.
4. Como lo que estudiamos es lo que ocurre en una determinada población en un momento determinado no podemos determinar si la exposición precede o es una consecuencia de la enfermedad.
5. Como no sabemos cuantas personas han pasado la enfermedad, no podemos distinguir entre factores de riesgo (causas) y factores pronóstico, factores relacionados con la probabilidad de que la enfermedad se haga crónica y pueda ser detectada. No proporciona estimadores de incidencia o de riesgo relativo.
6. Posible sesgo de supervivencia: los casos observados pueden tener una mayor supervivencia.

5. ANÁLISIS

Antes de empezar con el análisis hay que hacer notar que como en todo estudio observacional y analítico hay que prestar atención en los siguientes aspectos:

1. El cálculo del tamaño de la muestra
2. El análisis de las interacciones o de los factores modificadores del efecto
3. El control de los factores de confusión.

Sin embargo, al ser estos aspectos comunes a los estudios de cohortes y de casos y testigos, y para no repetir lo mismo expuestos ya en otros capítulos de este libro, no serán estudiados en este apartado.

5.1. Estimaciones puntuales.

	Enfermo	Sano	
Expuesto	a	b	N_1
No Expuesto	c	d	N_0
	M_1	M_0	N

Como hemos comentado los estudios de prevalencia tienen características comunes con los estudios de cohortes y de casos y testigos, por lo tanto podemos analizarlos utilizando ambos métodos.

a) Aproximación de un estudio transversal a un estudio de cohortes de incidencias acumuladas

Razón de proporciones (RP)

$$RP = \frac{P_1}{P_0} = \frac{a/N_1}{c/N_0}$$

P_1: Prevalencia en expuestos

P_0: Prevalencia en no expuestos

$$\text{Varianza}(\ln RP) = \frac{1}{a} + \frac{1}{c} - \frac{1}{N_1} - \frac{1}{N_0}$$

¿Cuándo podemos utilizar esta aproximación?

1. Cuando la enfermedad no influye en el hecho de estar en el estudio.
2. Cuando la enfermedad no influye en el hecho de estar expuesto.

b) Aproximación de un estudio transversal a un estudio de casos y controles

Población estable en estado de equilibrio.

$$OR = \frac{ad}{cb}$$

Prevalencia de exposición en enfermos $= \dfrac{a}{a+c}$

Prevalencia de exposición en sanos $= \dfrac{b}{b+d}$

$$\text{Varianza}(\ln OR) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

c) Relación entre *RP* y *OR*:

$$OR = RP\frac{1-P_{x-}}{1-P_{x+}}$$

P_{x-} = Prevalencia de la enfermedad en el grupo no expuesto; c / N_0

P_{x+} = Prevalencia de la enfermedad en el grupo de expuesto; a / N_1

Si la prevalencia de la enfermedad es baja (menor del 20%), $1\text{-}P_{x-}$ y $1\text{-}P_{x+}$ valdrán prácticamente 1, y *OR* será muy similar a *RP*. Por tanto, para enfermedades poco frecuentes *OR* será prácticamente igual a *RP*.

5.2. Estimación por intervalos.

Se pueden utilizar cualquiera de los métodos utilizados en los análisis de cohortes o de casos y testigos. Ponemos como ejemplo, el método de las series de Taylor. Este método es idéntico para los estudios de casos y controles, sólo se debe sustituir RR por la OR

$$IC_{95\%} = RR \cdot \exp\left(\pm 1{,}96\sqrt{\text{varianza } RR}\right)$$

5.3. Medidas de impacto

Son las mismas que se utilizan en los estudios analíticos. Así, cuando el RR o la OR es mayor de 1 tenemos: la Proporción Atribuible en Expuestos (PAE) y la Proporción Atribuible en la Población (PAP).

$$PAE = \frac{RR-1}{RR}$$

$$PAP = \frac{a}{M_1}\left(\frac{RR-1}{RR}\right)$$

Casos atribuidos a la exposición = PAE x a = PAP x M_1

Si la *OR* o el *RR* es menor de 1 tenemos: Proporción Prevenida por la Exposición (*PPE*) y la Proporción Prevenida en la Población (*PPP*).

$$PPE = 1 - RR$$

$$PPP = \frac{N_1}{N}(1-RR)$$

Casos evitados por la exposición = PPE x a / $(1\text{-}PPE)$

Se pueden calcular los intervalos de confianza de estas medidas a partir de los obtenidos con la OR y el RR.

5.4. Prueba de significación estadística. Igual con la *OR*

$H_0: RR = 1 \qquad H_1: RR \neq 1$

Tenemos que la raíz cuadrada de la ji-cuadrado de Mantel-Haenszel es igual a

$$\frac{a - \dfrac{M_1 N_1}{N}}{\sqrt{\dfrac{M_1 M_0 N_1 N_0}{N^2 (N-1)}}} = \frac{ad - bc}{\sqrt{\dfrac{M_1 M_0 N_1 N_0}{N-1}}}$$

Si N-1 es sustituido por N, tenemos la χ^2 de Pearson. La raíz cuadrada de la χ^2 de M-H se distribuye según N(0 , 1) o de una t, dependiendo del tamaño de la muestra, el valor de ´P´ se busca en una tabla de la normal; el valor de la ´P´ de la χ^2 se busca en la tabla de la χ^2 con un grado de libertad.

6. APLICACIONES

1. Cuando se desea estimar la prevalencia de una enfermedad para:
   - Distribuir recursos.
   - Saber qué probabilidad hay de que un paciente tenga una determinada enfermedad.

2. Para sugerir hipótesis etiológicas entre exposición y enfermedad que puedan ser confirmadas o rechazadas en estudios de cohortes o de casos o testigos

3. Para estimar la distribución de los valores de diferentes variables en diversos grupos de la población. Por ejemplo, la información obtenida en la Encuesta Nacional de Salud de España.

4. En enfermedades de larga duración e inicio lento es difícil saber cuándo empieza. Por ejemplo, bronquitis, enfermedades mentales, etc. En estos casos el diseño más útil es el de prevalencia

5. Cuando no se pueden realizar estudios de incidencia. Por ejemplo, el estudio de las malformaciones congénitas. Estos estudios son de prevalencia, sobre todo si se estudian al nacimiento, ya que sólo vemos aquellas malformaciones que sobreviven al embarazo. Para que fuera un estudio de incidencia además de estudiar éstas deberíamos estudias aquellas que son causas de un aborto terapéutico y las que provocan un aborto espontáneo.

6. Como primera etapa de un estudio de cohortes, ya que nuestro objetivo en este caso es seleccionar a personas sanas.

7. Estudios de Seroprevalencia. Si la prevalencia de un marcador serológico se estima en momentos diferentes, se obtiene una medida de los cambios temporales del estado del mismo. Estos estudios sirven para estudiar los niveles de protección y la cobertura de las vacunas.

7. COMUNICACIÓN DE LOS RESULTADOS DE LOS ESTUDIOS TRANSVERSALES. LA DECLARACIÓN STROBE (Strengthening the Reporting of Observational studies in Epidemiology)

La Declaración STROBE está constituida por una lista de puntos a tener en cuenta en la comunicación de los estudios realizados con los 3 diseños más importantes de la epidemiología analítica observacional: los estudios de cohortes, los estudios de casos y controles y los estudios

transversales. La única intención de esta declaración es ofrecer una guía sobre la forma adecuada de comunicar los estudios de investigación observacionales: estas recomendaciones no representan una «receta» para el diseño o la realización de los estudios. La Declaración STROBE fue desarrollada para ayudar a los autores a redactar sus estudios observacionales analíticos, para ayudar a los editores y los revisores que consideran la posible publicación de estos artículos, y para ayudar a los lectores que evalúan de forma crítica los artículos publicados. A continuación aparece la lista referida las estudios transversales. La lista está en inglés ya que a la fecha de redacción de este capítulo (septiembre 2009) era la única versión disponible.

STROBE Statement—Checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram

Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
Outcome data	15*	Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

**Give information separately for exposed and unexposed groups.*

Note: *An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.*

8. BIBLIOGRAFÍA

ARGIMÓN PALLAS JM, JIMÉNEZ VILLA J. *Métodos de investigación clínica y epidemiológica.* 2ª edición: Harcourt.. Madrid 2000.

KLEINBAUM DG, KUPPER LL, MORGESTERN H. *Epidemiologic research: principles and quantitative methods.* Van Nostrand Reinhold Company. New York . 1982

Ministerio de Sanidad y Política Social. Encuesta Nacional de salud de España. http://www.msps.es/estadEstudios/estadisticas/encuestaNacional/ense.htm

NORELL S. *Diseño de estudios epidemiológicos:* Madrid. Siglo XXI de España editores. 1994.

ROCA ANTONIO J, MUÑOZ VILLEGAS A. *Los estudios de prevalencia.* En: Martínez Navarro F, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V. Salud Pública: McGraw Hill Interamericana. Madrid 1998. Capítulo 11 pp.179-198.

SCHIAFFINO A, RODRÍGUEZ M, PASARÍN MI, REGIDOR E, BORRELL C, FERNÁNDEZ E. *¿Odds ratio o razón de proporciones?. Su utilización en estudios transversales.* Gac Sanit 2003; 17(1): 70-4

Declaración STROBE *(Strengthening the Reporting of Observational studies in Epidemiology)* http://www.strobe-statement.org/

VAQUÉ J (EDITOR). *Prevalencia de las infecciones nosocomiales en los hospitales españoles.* Proyecto EPINE. http://www.sempsph.com/sempsph/index.php?option=com_content&view=article&id=267:epine-2009&catid=1:general&Itemid=10

IX. LOS ESTUDIOS ECOLÓGICOS

Gonzalo López-Abente

1. INTRODUCCIÓN

En el capítulo 3 se han revisado los diferentes tipos de estudios epidemiológicos, que en su mayor parte tienen sus raíces en los conceptos de la experimentación científica. Cuando los experimentos en epidemiología (ensayos clínicos, ensayos de intervención comunitaria, etc) no son factibles, los epidemiólogos diseñan estudios denominados 'observacionales' (estudios de cohortes, de casos y controles, transversales) que intentan simular lo que se habría obtenido si se hubiese realizado un experimento. Existe un tipo de estudio en el que la unidad de observación son grupos de individuos, los estudios ecológicos, que van a ser el objeto de este capítulo.

Los estudios ecológicos se diferencian de un estudio observacional convencional en que los niveles de exposición individual no se miden, o bien la información sobre la exposición, si se mide, no se vincula con la frecuencia de la enfermedad a nivel individual. La unidad típica de análisis estadístico es un área geográfica, como puede ser una sección censal, una provincia o un país. Para cada grupo o área, es posible estimar la distribución de exposiciones o al menos el nivel promedio de exposición, y podemos estimar la tasa de enfermedad general, pero carecemos de la información conjunta de medidas de niveles de exposición y estatus de enfermedad a nivel de individuos. Por ello, es imposible obtener la tasa de incidencia en los grupos expuestos y no expuestos a partir de datos ecológicos, debiendo recurrir a estimaciones indirectas. La estimación indirecta de efectos en los estudios ecológicos y temas claves como el control de variables de confusión, contienen complicaciones metodológicas que hacen de los mismos un área de la epidemiología con una alta especialización.

La necesidad de abordar este tipo de estudios surge de la dificultad de obtener datos de alta calidad sobre exposiciones ambientales y las variables relacionadas con ellas. Por este motivo, los estudios ecológicos han sido y son utilizados frecuentemente en diversas áreas de investigación. Hace años los estudios ecológicos se encuadraban bajo el epígrafe de simples estudios descriptivos en los que el análisis de las tasas de enfermedad se estratificaban por lugar y tiempo para explorar diferentes hipótesis prestando escasa atención a los métodos estadísticos de inferencia.

En la tabla 9-1 se muestra un ranking de los diferentes tipos de estudios ordenados según el grado de validez que se acepta en al ámbito epidemiológico, aunque según el contenido del estudio el orden podría ser diferente. En general los estudios ecológicos son colocados en la parte más baja de la escala. Sin embargo, los diseños de los estudios ecológicos puden incluir componentes que los hagan ascender varios niveles y son los diseños más utilizados para orientar los proyectos en muchos ámbitos.

Tabla 9-1. Validez de los diferentes diseños para la inferencia etiológica.

Validez	Tipos de diseños
La más alta	Ensayo clínico aleatorizado
	Estudio de cohortes prospectivo
	Estudio de cohortes retrospectivo
	Estudio caso-control anidado
	Análisis de series temporales
	Estudios de sección-transversa
	Estudio ecológico
	Análisis de agregaciones
	Estudio de casos
La más baja	Anécdota

Tomado de Künzli y Tager. Environ Health Perspect 1997; 105:1078-83

Los objetivos de este capítulo son: 1) despertar el interés del lector en los estudios ecológicos, 2) revisar su metodología insistiendo en su diseño y en la interpretación de los resultados y 3) proporcionar información sobre el camino que está tomando el desarrollo de los nuevos métodos en este área.

Para ilustrar los contenidos teóricos de este tema nos apoyaremos en los siguientes ejemplos:

1. *Instalaciones nucleares y cáncer.* Las instalaciones de producción de energía nuclear y las de obtención de su combustible producen una contaminación en su entorno que podría ser origen de diversas patologías en la población que vive en su proximidad. Inicialmente no se dispone de buenas mediciones o estimaciones de los niveles de exposición de estas poblaciones derivados del funcionamiento de la instalación. No obstante, se dispone de información individual sobre las defunciones en los municipios de su entorno. En esta situación se puede utilizar la distancia del municipio a la instalación como "indicador" de exposición y obtener una estimación del riesgo relativo de muerte en función de la distancia.

2. *Tipo de cultivos e incidencia de cáncer de encéfalo.* La provincia de Navarra muestra una mayor mortalidad e incidencia de cáncer de encéfalo que otras provincias españolas. Se sabe que los agricultores en España (al igual que en otros paises) tienen mayor riesgo de morir por cáncer de encéfalo que otras ocupaciones. Se cree que ello es debido a la exposición a diferentes sustancias químicas (productos fitosanitarios). ¿Podrían estar los tipos de cultivo presentes en Navarra asociados con la mayor incidencia de este tumor?.

2. CONCEPTOS BÁSICOS DE LOS ESTUDIOS ECOLOGICOS

Se define como ecológico, todo estudio en el que las unidades de análisis son poblaciones o grupos de personas más que individuos (Last).Es habitual considerar que si la variable de exposición central en un estudio epidemiológico, cualquiera que sea su diseño, está referida a un grupo de individuos, el estudio se define como ecológico. Sin embargo, hay quien define

como estudio ecológico en sentido estricto, sólo aquél en el que tanto la variable resultado ('outcome') como la exposición, son medidas agregadas.

2.1. Mediciones ecológicas.

El diseño y las posibilidades analíticas de los estudios ecológicos vienen determinadas por la disponibilidad de información y/o por la dificultad de obtener datos a nivel individual.

Ejemplo: Se quiere determinar la influencia de la contaminación atmosférica en la mortalidad por patología respiratoria en diferentes distritos de una ciudad. Las defunciones y el lugar de residencia se obtienen del registro de mortalidad (información individual) y los niveles de contaminantes de las estaciones de vigilancia ambiental (información referida a un área).

Lo que define a un estudio como ecológico es que incluya observaciones sobre grupos, lugares u organizaciones. El ejemplo anterior incluye variables sobre los individuos (edad, sexo, etc) y variables (ecológicas) sobre la exposición referidas a un área.

Las mediciones ecológicas pueden clasificarse en 3 tipos:

1. *Medidas agregadas.* Se trata de resumir las observaciones sobre individuos en cada grupo expresándolas como medias o proporciones. Por ejemplo la proporción de fumadores, de agricultores o de parados.
2. *Mediciones ambientales.* Son características generalmente del lugar en el que los miembros del grupo viven o trabajan (ejemplo: niveles de contaminación ambiental). Cada medición ambiental tiene su análogo a nivel individual y la exposición individual puede variar dentro del grupo. A esta categoría pertenecería las obtenidas de las matrices de exposición-ocupación (ver capítulo de epidemiología ocupacional) o la variable referida al tipo de cultivo en el ejemplo del cáncer de encéfalo en Navarra.
3. *Mediciones globales*. Son características de grupos, organizaciones o áreas para las que no existe una analogía con el nivel individual (al contrario que las mediciones ambientales). (ejemplo: densidad de población, tipo de sistema de salud, orientación del partido gobernante)

2.2. Niveles de análisis

La unidad de análisis es el nivel común al que se reducen todas las variables del estudio. Por ejemplo, en el estudio comentado al principio sobre la incidencia de cáncer de encéfalo y usos del suelo, la unidad de análisis es el municipio. En los estudios que se analizan a nivel individual se dispone de información de exposición para cada sujeto, pero en epidemiología ambiental es habitual que algunas de las variables explicativas sean medidas ecológicas.

En un *análisis ecológico completo,* todas las variables son medidas ecológicas (exposición o exposiciones, enfermedad u otras variables incluidas), ya que la unidad de análisis es el grupo. Ello implica que desconocemos la distribución conjunta de cualquier combinación de variables a nivel individual.

En un *análisis ecológico parcial* de tres o más variables puede tenerse información de la distribución conjunta de alguna de las variables en cada grupo. Por ejemplo, en el estudio de incidencia de cáncer de encéfalo se conocen la edad y el sexo de los casos (información individual) pero la exposición derivada de la residencia en un área concreta (municipio) es información ecológica.

El *análisis multinivel* es un tipo especial de modelización que combina el análisis efectuado a dos o más niveles. Ejemplo: la modelización de la incidencia de cáncer de encéfalo incluyendo el sexo y la edad como variables explicativas, además de variables socio-demográficas

(proporción de parados, analfabetos), variables de exposición como el tipo de cultivo y variables de localización espacial de cada municipio.

2.3. Niveles de Inferencia

Para distinguir entre inferencia biológica y ecológica usaremos un ejemplo: Hacemos un estudio sobre la eficacia en la prevención de accidentes mortales del uso de casco en motoristas. Si el objetivo es estimar el efecto individual de utilizar el casco la *inferencia causal es biológica* (efectos en el riesgo individual). Si el objetivo es cuantificar la tasa de mortalidad entre los motoristas en países con y sin obligatoriedad de uso de casco, la *inferencia es ecológica* (efectos en las tasas del grupo). Ha de tenerse en cuenta que la magnitud del *efecto ecológico* (efectos en el riesgo, o tasa, del grupo) depende, no sólo del *efecto biológico* (efectos en el riesgo individual) del uso del casco sino del grado de cumplimiento de la legislación en cada país. Además, la validez de la estimación del efecto ecológico dependerá de nuestra habilidad para controlar las diferencias entre estados en la distribución conjunta de variables de confusión, incluyendo variables a nivel individual como la edad y magnitud del uso de motos. Por último, el *efecto contextual* valora la influencia sobre el riesgo individual de una exposición ecológica.

2.4. Justificación de los estudios ecológicos

Existen numerosas razones que justifican la necesidad o conveniencia de llevar a cabo estudios ecológicos. A continuación señalamos algunas de las más destacadas:

1. *Bajo coste.* Utilización de diferentes fuentes de datos secundarias, que se enlazan para hacer el estudio de forma agregada. Por ejemplo, los datos de mortalidad de los municipios se juntan con los datos censales y de otras encuestas.
2. *Limitaciones en las mediciones de los estudios individuales.* En algunas situaciones y sobre todo en epidemiología ambiental, los estudios ecológicos son los únicos posibles. Ejemplo: El riesgo de cáncer en el entorno de focos contaminantes. Es muy difícil disponer de mediciones de exposición individuales. Los indicadores 'ecológicos' de exposición, como por ejemplo la distancia, son los únicos utilizables en muchas circunstancias.
3. *Limitaciones en el diseño de los estudios individuales.* Si una exposición tiene una escasa variabilidad en el área de estudio, un estudio individual no tiene interés práctico. Diseñar un estudio ecológico comparando áreas puede ser una solución.
4. *Interés de los efectos ecológicos.* Ejemplo: tratar de entender las diferencias en las tasas de enfermedad entre dos poblaciones, supone apuntar hacia un objetivo de inferencia ecológica. Esta es la situación habitual en la evaluación del impacto de procesos sociales o de políticas de intervención (ej.: legislación o programas de prevención).
5. *Simplicidad de análisis y presentación de resultados.* En la explotación de grandes encuestas periódicas (ej.: encuestas de salud) es habitual que, aunque los datos han sido obtenidos con cuestionarios individuales, el análisis y/o resultados se muestren de forma agregada, con presentación por años, provincias o regiones como unidad de análisis.

3. DISEÑO DE ESTUDIOS ECOLÓGICOS

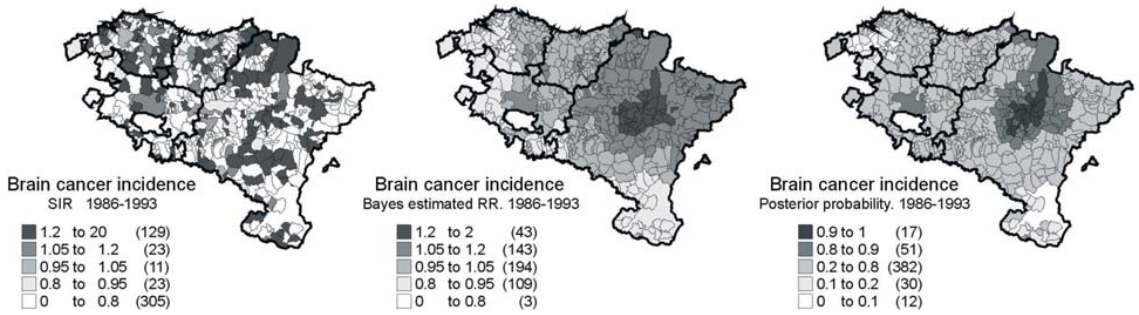
Los estudios ecológicos pueden clasificarse por: 1) el método de medición de la exposición (exploratorios y analíticos) y 2) el método de agrupamiento utilizado (diseños multi-grupo, tendencias temporales, diseños mixtos).

3.1. Diseños multi-grupo

Estudios exploratorios multi-grupo. Comparación de tasas de enfermedad entre muchas regiones en un mismo periodo. Dentro de esta categoría se incluirían todos los estudios de patrones espaciales que tratan de sugerir hipótesis centradas generalmente en etiologías ambientales. Un ejemplo podría ser el Atlas de mortalidad por cáncer en España, 1976-1992 (http://www2.uca.es/hospital/atlas92/www/introdu.html). Este atlas muestra la distribución provincial de la mortalidad por diferentes tumores malignos y otras causas. Muchos tumores malignos muestran un patrón geográfico fácilmente identificable. Los tumores que comparten factores de riesgo muestran una distribución similar (cáncer de pulmón, esófago, vejiga, laringe). El cáncer de estómago muestra un patrón único y muy diferente del resto de localizaciones tumorales. La existencia de 'patrones' apunta hacia determinantes ambientales de la enfermedad. Hay otras causas de muerte que muestran una escasa variabilidad geográfica (no muestran un patrón), como las leucemias o el cáncer de mama en mujeres. Ello puede orientar hacia el menor peso de factores de riesgo ambientales aunque en el caso del cáncer de mama puede indicar una distribución amplia y homogénea de ellos.

En los estudios geográficos, la simple comparación de tasas entre áreas a menudo se ve complicada por dos problemas estadísticos. El primero es que cuando se estudian enfermedades poco frecuentes o las áreas elegidas son muy pequeñas, el número de casos es muy bajo, y los datos muestran una gran variabilidad en la tasa estimada. De hecho las tasas más extremas las presentarán las áreas con menos población y con menos casos. El segundo problema es que las áreas más próximas tienden a tener tasas más parecidas que áreas distantes, porque factores de riesgo no medidos o no identificados tienden a agruparse en el espacio. Esto supone la presencia de fenómenos de autocorrelación. Afortunadamente ya se dispone de métodos estadísticos que aportan soluciones a ambos problemas, ajustando modelos espaciales autorregresivos a los datos y utilizando técnicas de suavización que permiten revelar los patrones presentes en los mapas. En la Figura 9-1 podemos ver un ejemplo de esto último. En el mapa de la izquierda se representa la Razón de Incidencia Estandarizada de cáncer de encéfalo por municipios en Navarra y Comunidad Autónoma Vasca. Dada la gran variabilidad aleatoria de los datos es muy difícil apreciar algún patrón. El mapa del medio muestra los riesgos relativos municipales estimados utilizando un modelo autorregresivo espacial y apunta a que las áreas de mayor riesgo se agrupan en el interior de Navarra.

Figura 9-1. Incidencia de cáncer de encéfalo por municipios en Navarra y Comunidad Autónoma Vasca, 1986-93. Razones de Incidencia Estandarizada (izquierda, mapa sin suavizar), riesgos relativos estimados (centro, mapa suavizado), y probabilidades posteriores de RR>1 (derecha). El número de municipios en cada categoría entre paréntesis)



Estudios analíticos multi-grupo. Estudio de la asociación entre un nivel de exposición medio y la prevalencia o la tasa de una enfermedad entre muchos grupos. Este tipo de estudios supone una extensión del diseño anterior mediante la inclusión de variables explicativas en los modelos. Es el diseño más común de estudio ecológico. El ejemplo que hemos comentado al principio del capítulo sobre instalaciones nucleares y mortalidad por cáncer contiene elementos

analíticos, al evaluar el efecto de la distancia a la instalación nuclear en el riesgo de muerte por cáncer. En el ejemplo sobre la incidencia de cáncer de encéfalo en Navarra la inclusión de la variable 'uso del suelo' y la estimación del riesgo relativo asociado a ella también es un elemento que permite calificarlo como analítico.

3.2. Diseños de tendencias temporales

Estudios exploratorios de tendencias temporales. Comparación de tasas en el tiempo en una población definida geográficamente. Los análisis edad-periodo-cohorte, tan utilizados en diferentes ámbitos, constituyen un diseño de este tipo. Básicamente consisten en estimar el peso de los tres componentes en la tendencia temporal observada. Cada uno de los componentes permite una interpretación específica y de utilidad en la formulación de hipótesis.

La monografía sobre tendencias de la mortalidad en España editada por el ISCIII, accesible en http://cne.isciii.es/cancer/publi3.htm, representa un ejemplo ilustrativo de este tipo de diseño.

Estudios analíticos de tendencias temporales. La inclusión de niveles de exposición medios u otras variables explicativas en un diseño de tendencias temporales lo convertirían en un estudio analítico.

3.3. Diseños mixtos

Morgenstern establece una tercera categoría de diseño que incluye ambos componentes de los dos anteriores, el espacial y el temporal. Los denomina *diseños mixtos.* En definitiva se trata de estudios que combinan ambas informaciones y estos modelos tratan de mostrar si una asociación que se observa en un estudio multigrupo se mantiene en el tiempo, o viceversa, si un patrón de tendencia temporal se mantiene entre diferentes grupos. A esta categoría pertenecerían todos los estudios de tendencias temporales en los que se introduce un componente espacial o los estudios de grupos múltiples en los que se explora la evolución temporal de los patrones. La inclusión de variables explicativas relacionadas con diferentes exposiciones convertirían el diseño de exploratorio en analítico.

3.4. Estimación del Efecto

En estudios con datos individuales los efectos se determinan comparando las tasas o el riesgo de enfermar (o morir) entre grupos expuestos y no expuestos. En los estudios ecológicos multigrupo no se pueden estimar los efectos de cada exposición directamente, ya que carecemos de la información de la distribución conjunta dentro de los grupos. La forma más habitual de análisis es utilizar modelos de regresión lineal o modelos log-lineales. Estos modelos permiten estimar el efecto para cada grupo y con un significado distinto dependiendo del tipo de modelo.

Los modelos lineales tienen la siguiente forma

$$Y = B_0 + B_1 X$$

Sea Y la tasa de enfermedad y X la exposición en los grupos. La tasa predicha ($Y_{x=1}$) en el grupo expuesto es $B_0 + B_1(1) = B_0 + B_1$ y la tasa en el grupo no-expuesto ($Y_{x=0}$) es $B_0 + B_1(0) = B_0$. La diferencia de tasas será $B_0 + B_1 - B_0 = B_1$ y la razón de tasas será $(B_0 + B_1)/B_0 = 1 + B_1/B_0$.

Los modelos log-lineales tienen la siguiente forma

$$\ln[Y] = B_0 + B_1 X$$

en estos modelos la razón de tasas es $\exp[B_1]$. La ventaja de estos modelos es que la variable dependiente no puede tomar valores negativos, mientras que en el modelo lineal si. Es por ello que en la actualidad son los más utilizados.

3.5. Variables de confusión y modificadoras del efecto

El tratamiento de las posibles variables de confusión en estos modelos es el habitual. Las variables ecológicas han de incluirse en el modelo como covariables para obtener una estimación ajustada del efecto en cuestión. Otra posibilidad de tratamiento de las variables de confusión es la obtención de tasas estandarizadas para esas variables. Así, en el ejemplo de la incidencia del cáncer de encéfalo lo que se modeliza es la RIE (Razón de Incidencia Estandarizada). Ello supone que la variable edad ya se tiene en cuenta en el modelo mediante el proceso de estandarización. Sin embargo, la reducción del sesgo inducido por las variables de confusión no será completa si el resto de las variables predictoras no se ajustan también por las mismas variables (Greenland 1992). En la práctica esto último es muy difícil de hacer ya que requiere conocer la distribución conjunta de las covariables y la exposición dentro de los grupos, lo que suele ser inviable en los estudios ecológicos.

4. PROBLEMAS METODOLÓGICOS

La inferencia causal en estos estudios se encuentra seriamente limitada por problemas metodológicos que van a limitar especialmente la inferencia la biológica.

4.1. El sesgo ecológico (falacia ecológica)

El sesgo ecológico es el error que se produce cuando la estimación del efecto ecológico esperado no coincide con el verdadero valor del efecto biológico a nivel individual. Hemos comentado que en los estudio ecológicos carecemos de información sobre la distribución conjunta de las variables. La falacia ecológica consiste en atribuir a todos los miembros del grupo las características que no poseen los individuos.

Un ejemplo muy utilizado para ilustrar este fenómeno es el del incremento de la tasa de suicidio en función de la religión (Durkheim 1951). Ese estudio muestra que el riesgo relativo es cercano a 8 en la región predominantemente protestante comparado con la no protestante. Sin embargo, desconocemos la religión de los que se han suicidado. Podría suceder que los que se han suicidado en la región con predominio protestante fuesen precisamente los católicos. Otro ejemplo nos muestra que a pesar de la gran correlación existente entre la mortalidad por cáncer de mama y la ingesta per capita de grasa por países, no está claro que las mujeres que han fallecido por cáncer de mama tengan una mayor ingesta de grasas a nivel individual.

El sesgo ecológico surge a partir de tres fuentes cuando se utiliza un análisis mediante regresión lineal para estimar el efecto crudo:

1. *Sesgo intra-grupo.* El sesgo ecológico puede ser originado por la presencia de sesgos intra-grupo debido a confusión, selección o información, aunque los efectos intra-grupo no se estimen. Ejemplo: si existe en cada grupo un efecto de confusión será de esperar que la estimación ecológica del efecto también esté sesgada.
2. *Confusión por grupo.* Si la tasa de enfermedad basal en la población no expuesta varía en los grupos, especialmente si la correlación ecológica entre el nivel de exposición medio y la tasa basal es distinta de cero, se producirá un sesgo ecológico.
3. *Modificación del efecto por grupo (escala aditiva).* Puede producirse un sesgo ecológico si la diferencia de tasas para el efecto de la exposición a nivel individual varía en los grupos.

Para ver en la práctica las consecuencias del sesgo ecológico es útil comparar los resultados de un análisis con datos individuales y con datos agregados. En la tabla 9-2 se muestra uno de estos ejemplos en el que existe un fuerte sesgo ecológico debido a la heterogeneidad de la

diferencia de tasas entre los grupos (modificación del efecto por el grupo). Nótese que la tasa en no expuestos es idéntica en los tres grupos por lo que no hay efecto de confusión por grupo.

Tabla 9-2. Número de casos nuevos, personas-año de seguimiento y tasa de enfermedad (Y por 100.000/años) por grupo y status de exposición (x); distribución para cada grupo y resultados del análisis individual y ecológico. Datos hipotéticos que muestran el sesgo ecológico debido a la modificación del efecto por grupo [Morgenstern 1998]

Exposición	Grupo 1			Grupo 2			Grupo 3		
	Casos	P-año	Tasa	Casos	P-año	Tasa	Casos	P-año	Tasa
Expuestos	20	7000	286	20	20000	200	20	13000	154
No-expuestos	13	13000	100	10	10000	100	7	7000	100
Total	33	20000	165	30	20000	150	27	20000	135
% expuestos			35			50			65
Diferencia de tasas ($x10^5$)			186			100			54
Razón de tasas			2,9			2,0			1,5

Análisis individual	Análisis ecológico: modelo lineal
Razon tasas cruda = 2,0	Y = 200 -100X
Razon tasas ajustada= 2,0	Razón de tasas = 0,50

En la siguiente tabla se muestra un ejemplo sintético de inexistencia de sesgo ecológico. Al ser la diferencia de tasas uniforme entre grupos (aunque la razón de tasas varía) y no existir correlación entre el porcentaje de expuestos y la tasa en no expuestos, no se produce el sesgo ecológico coincidiendo la estimación de efecto individual y ecológica.

Tabla 9-3. Número de casos nuevos, personas-año de seguimiento y tasa de enfermedad (Y por 100.000/años) por grupo y status de exposición (x); distribución para cada grupo y resultados del análisis individual y ecológico. Datos hipotéticos que muestran la inexistencia de sesgo ecológico [Morgenstern 1998]

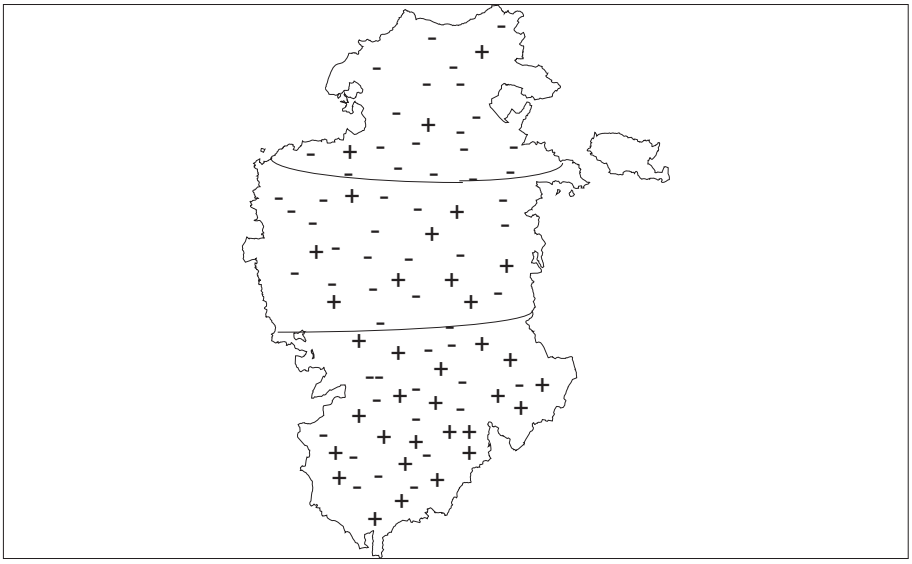
Exposición	Grupo 1			Grupo 2			Grupo 3		
	Casos	P-año	Tasa	Casos	P-año	Tasa	Casos	P-año	Tasa
Expuestos	16	8000	200	30	10000	300	24	12000	200
No-expuestos	12	12000	100	20	10000	200	8	8000	100
Total	28	20000	140	50	20000	250	32	20000	160
% expuestos			40			50			60
Diferencia de tasas ($x10^5$)			100			100			100
Razón de tasas			2,0			1,5			2,0

Análisis individual	Análisis ecológico: modelo lineal
Razon tasas cruda = 1,8	Y = 133 + 100X
Razon tasas ajustada= 1,8	Razón de tasas = 1,8

Desafortunadamente es imposible comprobar la validez de la estimación efectuada en un estudio ecológico porque las condiciones de validez se definen en términos de asociación a nivel individual. Esta es una seria limitación de los estudios ecológicos para hacer inferencia biológica. Además, al ajustar un modelo de regresión ecológico no hay ningún parámetro que nos indique la presencia, dirección y magnitud del sesgo. Un modelo puede ofrecer un excelente ajuste a los datos y estar sesgado y otro puede tener peor ajuste y no estarlo. Si se han hecho análisis en paralelo de datos individuales y ecológicos. Estos estudios muestran que la inclusión de componentes espaciales en los modelos acercan ambas estimaciones lo que da pie al siguiente apartado.

Efecto de confusión debido a la localización (Bernardinelli & Clayton 1992). Un amplio sector de los estudios ecológicos se centra en las comparaciones múltiples de áreas geográficas. Si el patrón de variación de una covariable es similar a la del riesgo de enfermar, la localización (geográfica) pude actuar como una variable de confusión. En la figura 9-2 podemos mostrar este fenómeno. En el mapa el signo (+) indica un área con alta exposición y (-) indica baja exposición. En el mapa hay una clara tendencia a una mayor frecuencia de exposición en el sur. Si existiese un gradiente norte-sur en el riesgo de enfermar, podría existir un efecto de confusión por localización, o de forma más precisa, por alguna influencia no medida responsable de la variación geográfica de la enfermedad. Este efecto de confusión podría ser minimizado estratificando el mapa en regiones en las que haya menor variación geográfica tanto en la exposición como en las tasas. El efecto (ajustado) se estima por comparaciones locales de areas expuestas y no expuestas dentro de cada region (exactamente igual que en un análisis estratificado). Sin embargo, aunque este análisis tiene la virtud de la simplicidad, en la práctica puede ser difícil de hacer. El marco de los modelos multinivel que comentaremos más adelante ofrece una solución técnica a este problema, mediante la inclusión de términos espaciales en los modelos. La inclusión de los términos espaciales (variable indicadora de cada área y de sus áreas vecinas) entre otras virtudes provoca la eliminación del efecto de confusión debido a la localización.

Figura 9-2. Efecto de confusión por localización



4.2. El control de variables de confusión

Las condiciones para que una covariable sea un factor de confusión son diferentes en los estudios individuales y los ecológicos. Mientras que a nivel individual un factor de riesgo debe

estar asociado con la exposición para ser una variable de confusión, en un estudio ecológico multigrupo un factor de riesgo, puede producir confusión aunque no esté asociado con la exposición en cada grupo, especialmente si el factor de riesgo está asociado ecológicamente con la exposición a través de los grupos. A la inversa, un factor de riesgo que es un factor de confusión dentro de los grupos puede no producir sesgo ecológico si no está asociado ecológicamente con la exposición a través de los grupos.

El control de los factores de confusión es más problemático en los estudios ecológicos que en los individuales. Morgenstern (1998) proporciona las condiciones suficientes para que no se produzca sesgo ecológico en un estudio de comparaciones múltiples. Sin embargo, disponiendo únicamente de los datos ecológicos es imposible comprobar si se cumplen esas condiciones, siendo impredecibles la naturaleza y magnitud del sesgo al hacer inferencias biológicas a partir de análisis ecológicos. Ante tantas dificultades Morgenstern proporciona un consejo general para el necesario control de las variables de confusión. La mejor aproximación para reducir el sesgo ecológico es incluir variables para categorías de su distribución conjunta dentro de los grupos. Por ejemplo, para controlar 'ecológicamente' por raza y sexo, lo convencional sería controlar por las proporciones de hombres (o mujeres) y blancos (o no-blancos), pero también podría controlarse por la proporción de mujeres blancas, hombres no-blancos y mujeres no-blancas (utilizando como referencia a los hombres blancos).

4.3. Errores de clasificación intra-grupo

Los principios de los errores de clasificación con los que los epidemiólogos estamos familiarizados y utilizamos en la interpretación de los resultados de los estudios observacionales no son aplicables a los estudios ecológicos. En los estudios individuales con errores de clasificación 'no-diferencial' sabemos que el sesgo induce a que la estimaciones se acerquen hacia la no existencia de efecto (ver capítulo 5). En los estudios ecológicos multi-grupo, esto no es así cuando la variable de exposición es una medida agregada. En los errores de clasificación no-diferencial de una exposición dicotómica afectando a diferentes grupos, la estimación del efecto puede alejarse de la hipótesis nula (Brenner 1992).

4.4. Estrategias para reducir el sesgo ecológico

Las principales estrategias que podemos utilizar para prevenir o mitigar los efectos del sesgo ecológico se relacionan a continuación:

1. Utilizar unidades geográficas lo más pequeñas posibles (mejor provincias que estados, mejor municipios que provincias). Esto hace las unidades de análisis más homogéneas respecto a la exposición aunque no siempre es posible por la disponibilidad de los datos.
2. Otro aspecto lógico en la reducción del sesgo sería estudiar unidades mutuamente comparables respecto a covariables, y si esto no es posible, ha de establecerse un sistema de control de las variables de confusión en el análisis.
3. Si existen algunas regiones con resultados poco usuales respecto al resultado, exposición o combinación de covariables, ha de examinarse el efecto de excluir esas regiones del análisis (análisis de influencia).
4. Inclusión de componentes espaciales en los modelos.

Greenland (1992) proporciona, además de las bases teóricas para comprender los diferentes tipos de sesgos, una guía para minimizar estos sesgos utilizando diseños más apropiados

5. MODELOS MULTINIVEL

En el ejemplo de la incidencia de cáncer en Navarra disponemos de información individual de los casos nuevos como es la edad, el sexo y el municipio de residencia. Imaginemos que disponemos de información individual del resto de la población respecto a estas mismas variables. Ello nos permite construir modelos de regresión para representar la variación del riesgo por municipio de residencia. La forma de modelizar esto más eficiente sería mediante modelos de regresión de Poisson con datos agregados. A esta información podemos añadir variables agregadas referidas a cada municipio como son la proporción de analfabetos, de parados y algún indicador de nivel socio-económico. El hecho de residir en un municipio supone la exposición a diferentes componentes contextuales como sufrir una determinada gestión municipal, respirar el mismo aire, beber la misma agua, tener unos hábitos de vida más parecidos o estar expuesto a factores ambientales específicos. Las existencia de factores contextuales como los citados pueden abarcar a diferentes municipios ya que muchos de ellos están muy próximos. Hay exposiciones que son compartidas por municipios que están próximos o municipios vecinos. Los municipios 'vecinos' presentan características ambientales que no son independientes lo que implica la existencia de autocorrelación.

En este ejemplo identificamos variables que pertenecen a diferentes 'jerarquías'. El nivel 1 serían los datos individuales (edad y sexo), el nivel 2 serían los datos agregados referidos a los municipios y un tercer nivel sería los datos contextuales y espaciales (ubicación del municipio y sus vecinos que constituyen un grupo). Reconocer que los distintos individuos pertenecen a distintos grupos tiene consecuencias importantes a la hora de analizar los datos. La similitud de los individuos (o de su entorno) pertenecientes a un grupo establece una estructura de correlación intracontextual que dificulta el cumplimiento de la hipótesis de independencia sobre la que están basados los modelos de regresión tradicionales. El esfuerzo hecho en los últimos años para adaptar esta estructura jerárquica de los datos al marco de los modelos lineales generalizados ha dado como resultado los llamados modelos multinivel o modelos jerárquicos (en la literatura también se conocen como modelos mixtos o modelos GLMM 'Generalised Linear Mixed Models'). La calificación de modelos mixtos implica la posibilidad de construir modelos que incluyen términos de efectos fijos y términos de efectos aleatorios.

En el ejemplo que nos ocupa, los datos referidos a la información espacial (cada municipio con sus vecinos), esto es, la información de 'grupo' ('clustering' en la literatura anglosajona) ha de ser introducida en el modelo como término de efectos aleatorios. Ello tiene considerables ventajas. Una de ellas es que la inclusión de estos términos de efectos aleatorios, cuando los datos muestran una distribución que no se corresponde exactamente con la distribución de probabilidad utilizada (sobredispersión), dan cuenta de los fenómenos de sobredispersión corrigiendo la varianza de los parámetros. Otra es la ya comentada superación de la hipótesis de independencia de las observaciones. Otra es el control del efecto de confusión por localización que comentamos en el apartado anterior por la inclusión del término espacial en el modelo. Y por último hay que comentar el 'filtro' que supone a la gran variabilidad aleatoria de las observaciones cuando se trabaja con áreas pequeñas, permitiendo obtener unas estimaciones 'suavizadas' del efecto por áreas.

En este apartado hemos hecho una mera introducción a la idoneidad de los modelos multinivel, ya que la explicación en profundidad de las técnicas supera el contexto de este libro. Actualmente la aplicación de estos modelos está facilitado por la existencia de diferentes programas de ordenador.

En resumen, los modelos multinivel ofrecen distintas ventajas sobre los modelos tradicionales y el precio a pagar es una mayor complejidad tanto del marco teórico como de los

modelos propuestos para analizar los datos y por tanto de la comunicación de los resultados. Sin embargo, en los estudios ecológicos, lo habitual es enfrentarse con estructuras jerárquicas, con lo que la demanda de utilización de estos modelos está en aumento.

6. NOTA

Para la redacción de este capítulo hemos seguido muy de cerca el texto de Hal Morgenstern ya que consideramos que contiene un notable esfuerzo de síntesis y sistematización en la explicación de este tipo de estudios (Morgenstern 1998).

7. BIBLIOGRAFÍA

BRENNER H, GREENLAND S, SAVITZ DA. *The effects of nondifferential confounder misclassification in ecologic studies.* Epidemiology. 1992 Sep;3(5):456-9.

CLAYTON DG, BERNARDINELLI L, MONTOMOLI C. *Spatial correlation in ecological analysis.* Int J Epidemiol. 1993 Dec;22(6):1193-202.

GOLDSTEIN, H., RASBASH, J., PLEWIS, I., DRAPER, D., ET AL. (1998). *A user's guide to MLwiN.* London, Institute of Education.

GREENLAND S. *Divergent biases in ecologic and individual-level studies.* Stat Med. 1992 Jun 30;11(9):1209-23.

GREENLAND S, MORGENSTERN H. *Ecological bias, confounding, and effect modification.* Int J Epidemiol. 1989 Mar;18(1):269-74. Review. Erratum in: Int J Epidemiol 1991 Sep;20(3):824.

KENNETH J. ROTHMAN. *Methodologic Frontiers in Environmental Epidemiology.* Environmental Health Perspectives Volume 101, Supplement 4, December 1993

KÜNZLI N, TAGER IB. *The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies.* Environ Health Perspect. 1997 Oct;105(10):1078-83. Review.

LAGARDE F, PERSHAGEN G. *Parallel analyses of individual and ecologic data on residential radon, cofactors, and lung cancer in Sweden.* Am J Epidemiol 149, 268-274, 1999.

LÓPEZ-ABENTE G, ARAGONÉS N, POLLÁN M. *Solid-tumor mortality in the vicinity of uranium cycle facilities and nuclear power plants in Spain.* Environ Health Perspect. 2001 Jul;109(7):721-9.

LÓPEZ-ABENTE G, POLLÁN M, ARDANAZ E, ERREZOLA M.*Geographical pattern of brain cancer incidence in the Navarre and Basque Country regions of Spain.* Occup Environ Med. 2003 Jul;60(7):504-508.

MORGENSTERN H. *Ecologic studies in modern epidemiology.* In: Modern Epidemiology (Rothman K, Greenland S, eds). 2nd ed. Lippincott Raven Publishers, 1998;459-480.

PINHEIRO JC AND BATES DM. *Mixed-Effects Models in S and S-PLUS.* Statistics and Computing Series, Springer-Verlag, New York, NY, 2000.

PRENTICE RL, KAKAR F, HURSTING S, SHEPPARD L, KLEIN R, KUSHI LH. *Aspects of the rationale for the Women's Health Trial.* J Natl Cancer Inst. 1988 Aug 3;80(11):802-14.

RICHARDSON S, STUCKER I, HEMON D. *Comparison of relative risks obtained in ecological and individual studies: some methodological considerations.* Int J Epidemiol 16, 111-120, 1987.

SÁNCHEZ-CANTALEJO E, OCAÑA-RIOLA R. *Los modelos multinivel o la importancia de la jerarquía.* Gaceta Sanitaria 13(5):391-398, 1999.

SPIEGELHALTER D, THOMAS D, BEST N, ET AL. *BUGS: Bayesian inference using Gibbs sampling. Version 0.50.* Cambridge, MRC. Biostatistics Unit. 1996.

X. ESTUDIOS EXPERIMENTALES

Miguel Ángel Royo Bordonada
José María Martín Moreno

La Real Academia de la Lengua Española establece que un experimento consiste en hacer operaciones destinadas a descubrir, comprobar o demostrar determinados fenómenos o principios científicos. Para la mayoría de los científicos, esta definición abarca cualquier conjunto de observaciones, realizadas bajo condiciones controladas, que permitan contrastar una hipótesis científica previamente planteada. Por ejemplo, Otto Von Guericke (1602-1686) propuso la teoría del giro de los cometas alrededor del sol en una órbita elíptica excéntrica y un año después de su muerte, Newton estableció la Ley de Gravitación Universal. En 1682, Halley descubrió el cometa que lleva su nombre y se propuso comprobar la hipótesis de Otto. A partir de las leyes de Newton calculó que el Halley recorría su órbita en 76 años, por lo que afirmó que era el mismo cometa que se había visto en ocasiones anteriores (cada 75-76 años antes) y postuló que se volvería a ver en 1758. Lamentablemente, Halley no viviría lo suficente como para poder contrastar su hipótesis personalmente. No obstante, Joham Georg Palitzch, un terrateniente alemán aficionado a la astronomía, estuvo observando el firmamento a la espera del cometa durante 1758. Por fin, el 25 de diciembre de 1758 observó el cometa Halley en el firmamento, confirmando la hipótesis científica planteada por Otto. Desde esta perspectiva, los diseños vistos en los capítulos precedentes (cohortes y casos-control) constituirían estudios experimentales. Sin embargo, en epidemiología el estudio experimental implica un aspecto adicional: el investigador manipula las condiciones del estudio, con el objeto de averiguar el efecto que tal manipulación (intervención) tiene sobre las observaciones realizadas. Por ejemplo, un investigador se dispone a analizar el posible efecto cancerígeno de las dioxinas y para ello suministra altas concentraciones de estas sustancias a un grupo de ratones a través de la alimentación.

Aunque el elemento central que define un estudio experimental sea la introducción de una intervención que altera las condiciones del mismo y es asignada a los participantes por el investigador, para que podamos hablar de un experimento válido desde el punto de vista científico, es preciso que además se cumplan las siguientes condiciones:

1. La única razón por la que los sujetos reciben la intervención bajo estudio es el cumplimiento del protocolo del estudio.
2. En el estudio existe una serie de sujetos, denominados grupo control, que no reciben la intervención cuyo efecto se desea analizar.
3. La asignación de la intervención a una serie de sujetos, grupo de intervención, se lleva a cabo por un mecanismo debido al azar.

Como explicaremos con más detalle a lo largo de este capítulo, estos requisitos representan condiciones “sine qua non” para descartar que los efectos observados puedan deberse a factores desconocidos o no controlados y, de esta forma, poder atribuir con seguridad esos resultados a la intervención bajo estudio. Para lograr este objetivo es necesario controlar todos los factores que pueden afectar de forma relevante la condición bajo estudio; es decir, necesitamos crear unas condiciones tales que el único factor que presente variación entre los grupos de comparación sea la intervención cuyo efecto pretendemos evaluar.

Por desgracia, en ciencias sociales y biomédicas el número de factores que afectan a la mayoría de las condiciones de interés es tan numeroso, complejo y en muchos casos oculto o

desconocido, que es imposible hacerlo uniforme. Por ejemplo, en la investigación de las causas del cáncer es imposible crear un conjunto de condiciones que conduzcan de forma invariable al desarrollo de un cáncer en un periodo fijado de tiempo, incluso aún cuando la población bajo estudio sea un grupo de ratones clonados en el laboratorio. Inevitablemente, siempre va a existir la denominada "variación biológica", es decir, la variación que se produce en el conjunto de factores que contribuyen a que se desarrolle la condición bajo estudio. Aunque consiguiéramos mantener las mismas condiciones de temperatura, humedad, hábitat e iluminación, lo cual es bastante poco probable, es imposible hacer que todos los animales coman exactamente la misma cantidad o realicen ejercicio físico con idéntica frecuencia e intensidad, por no hablar de diferencias individuales en el metabolismo basal. En resumidas cuentas, en la investigación epidemiológica la idea de crear un duplicado exacto de un conjunto de circunstancias, en el que sólo varía un factor relevante, no es realista. No obstante, puede afirmarse que un experimento es aceptable si la variación de los factores extraños es demasiado pequeña como para afectar la condición bajo estudio de forma importante en comparación con el efecto de la intervención bajo estudio. En el ejemplo de las causas del cáncer, si la alimentación de los ratones influye en el desarrollo del cáncer, lo cual parece bastante verosímil, el control inadecuado de la misma puede suponer un problema. Si la variación en la alimentación es muy pequeña, es mucho menos probable que pueda influir sobre los resultados del experimento de forma importante.

Por tanto, cuando los estudios experimentales son factibles, su diseño está guiado por principios que reducen la variación de los factores extraños en comparación con el factor de interés. Sin embargo, como ya se ha mencionado en capítulos precedentes, los estudios experimentales no siempre son factibles, por razones de tipo ético o logístico. En estos casos, los epidemiólogos diseñan estudios analíticos observacionales para simular lo que podría haberse obtenido si se hubiese llevado a cabo un estudio experimental. Es decir, los principios utilizados en el diseño de estudios observacionales son los mismos que en el caso de los experimentales: reducir la variación de los factores extraños en comparación con el factor de interés. Para ello se utilizan tanto técnicas de diseño del estudio como de análisis de los datos.

1. ESTUDIOS CUASI-EXPERIMENTALES

Para algunos, únicamente podemos hablar con propiedad de estudio experimental cuando existe un grupo control que no recibe la intervención bajo estudio y sirve como grupo de referencia para establecer las comparaciones, siempre que la asignación de los participantes en el estudio a uno u otro grupo se haya llevado a cabo mediante un mecanismo debido al azar. Sin embargo, existen muchas situaciones en las que nos encontramos con que el investigador manipula las condiciones del estudio (introduce una intervención) sin que se den las condiciones previamente mencionadas (Tabla 10-1). El problema con este tipo de estudios es que carecen del grado de validez interna que se precisa para atribuir el efecto observado a la intervención bajo estudio.

Tabla 10-1. Clasificación de los estudios experimentales.

- Estudios cuasi-experimentales
  - Estudios no controlados
  - Estudios con controles históricos
  - Estudios controlados no aleatorizados

- Estudios controlados y aleatorizados
  - Ensayos de prevención primaria
    - Ensayo de campo
    - Ensayo de intervención comunitaria
  - Ensayo clínico
    - De grupos paralelos
    - Cruzado
    - Factorial

1.1 Estudios no controlados

Un grupo control es un conjunto de personas que comparte con el grupo de sujetos tratados todos los predictores de las variables resultado excepto el tratamiento a evaluar. Por tanto, su papel consiste en controlar la acción de los otros predictores en el efecto de la intervención que se está ensayando. Es decir, el grupo control informa del valor de la variable resultado en ausencia de intervención.

Los estudios no controlados, también denominados estudios antes-después, comparan los resultados del efecto estudiado en el mismo grupo de individuos antes y después de recibir la intervención. En este tipo de investigación no existe ninguna garantía de que el efecto observado sea debido a la intervención bajo estudio, también puede ser un debido a un efecto placebo, a un efecto período, a la propia historia natural de la enfermedad o a fenómenos puramente estadísticos, como la regresión a la media. El efecto período se refiere a todos aquellos factores que pudiendo tener alguna influencia sobre la condición de interés (clima, hábitat, estilos de vida, etc.), se hayan modificado a lo largo del período de estudio. A modo de ejemplo, supongamos que se lleva a cabo un estudio en Madrid entre los meses de marzo y junio para comparar un nuevo fármaco diurético en la terapia antihipertensiva con el tratamiento diurético estándar. La mejora del clima durante el período de estudio se asocia con un incremento en la práctica de ejercicio físico, que a su vez tiene un efecto beneficioso sobre las cifras de la tensión arterial, efecto que podría atribuirse de forma indebida al nuevo fármaco. El efecto placebo consiste en que el sólo hecho de administrar una intervención produce un efecto sobre el individuo, independientemente del principio activo que contenga la intervención.

1.2. Estudios con controles históricos

En este tipo de estudio, se compara un grupo de pacientes que están recibiendo la intervención cuyo efecto se pretende determinar con pacientes que recibieron una intervención diferente en el pasado. En este caso, el problema es que no podemos garantizar la comparabilidad entre los dos grupos, es decir, no podemos asegurar que los sujetos de ambos grupos son similares en todo excepto en la intervención estudiada. Con el paso del tiempo, nos podemos encontrar con diferencias en los procedimientos de selección de los pacientes susceptibles de ser intervenidos (edad, estadío clínico o anatomo-patológico, gravedad, patologías asociadas, …), en los tratamientos auxiliares aplicados, en los criterios diagnósticos y de valoración de la

respuesta al tratamiento, etc. En estos casos, nada nos permite afirmar que el efecto observado sea debido a la intervención estudiada y no a cualquier otra característica en la que difieren los grupos comparados. En 1997 Chalmers y cols. realizaron una revisión de los ensayos publicados sobre el efecto de la terapia anticoagulante en la mortalidad tras infarto de miocadio y comprobaron que la reducción media de la tasa de mortalidad fue del 54% en los ensayos con controles históricos frente al 21% en los ensayos controlados y aleatorizados.

Lamentablemente, los estudios no controlados o con controles históricos tienden a exagerar el valor de un nuevo tratamiento, creando un estado de confusión que provoca que los médicos sean reacios a poner en marcha un ensayo clínico randomizado una vez que están (falsamente) convencidos de la eficacia del nuevo tratamiento.

1.3. Estudios controlados sin aleatorizar

En este tipo de estudios, la asignación de los sujetos participantes a los grupos de comparación se realiza siguiendo un mecanismo no aleatorio, que puede ser sistemático (en función del orden de entrada en el estudio o de la fecha de nacimiento) o simplemente a criterio del investigador. En ambos casos, el investigador conoce a priori la intervención que recibirá el individuo, lo que puede afectar tanto su decisión de incluirlo en el estudio como la de asignarlo a uno u otro grupo. La presencia de grupos de tamaño muy desequilibrado nos debería hacer sospechar de la existencia de intencionalidad a la hora de ubicar los sujetos en uno u otro grupo y de un posible sesgo de asignación asociado. En un estudio de prevención secundaria tras infarto de miocardio para evaluar la eficacia de una dieta determinada o de la práctica de ejercicio físico en la tasa de mortalidad, el investigador que confía en la bondad de esas intervenciones tenderá (de forma consciente o inconsciente) a asignar los pacientes que considera más motivados y cumplidores y menos graves al grupo de intervención. De forma similar, en un ensayo terapéutico para comparar el efecto de las versiones genérica y patentada de un fármaco, el médico encargado de asignar la intervención podría tener la tentación de dar a sus pacientes ricos a la versión patentada y a los pobres a la versión genérica. En ambos casos, el problema es que se han constituido dos grupos no comparables, ya que difieren no sólo en la intervención estudiada sino en muchas otras características conocidas o no (nivel socioeconómico, gravedad, hábitos de vida, predisposición genética, …) que podrían influir sobre el efecto investigado. Lo mismo puede ocurre si el investigador deja elegir al paciente entre las diferentes alternativas en función de sus propias preferencias. Por tanto, si no existe asignación aleatoria resulta difícil atribuir los resultados del estudio a la intervención, pudiendo deberse a diferencias en otros factores (conocidos o no) introducidas por el investigador en el momento de asignar la intervención.

Como ya hemos mencionado previamente, en los estudios experimentales propiamente dichos, también denominados ensayos, existe un grupo control y la intervención bajo estudio se asigna mediante un mecanismo debido al azar (aleatorio). Antes de pasar a analizar este tipo de estudios, es preciso describir los fundamentos de este mecanismo aleatorio de asignación, comúnmente denominado aleatorización, y la metodología desarrollada para su correcta aplicación.

2. ALEATORIZACIÓN

La aleatorización consiste en la asignación de las unidades experimentales (sujetos participantes) a dos o más intervenciones utilizando un mecanismo aleatorio de forma que ni el investigador ni el sujeto investigado conozcan la intervención que va a ser asignada en el momento de la inclusión de éste último en el estudio, una vez obtenido su consentimiento

informado para participar en el estudio. Aunque un estudio aleatorizado es necesariamente controlado, el grupo control podrá recibir una intervención activa, que habitualmente será la utilizada en condiciones normales para la condición bajo estudio, inactiva e inocua (placebo) o simplemente no recibir intervención alguna.

La aleatorización elimina el sesgo del investigador en la asignación de los sujetos, garantiza la validez de los niveles de significación de las pruebas estadísticas que se aplicarán en el análisis de los datos y tiende a producir grupos comparables respecto de características conocidas y desconocidas que pudieran afectar al desenlace de interés (factores pronósticos). Existen múltiples formas en las que se pueden producir sesgos en la asignación. A modo de ilustración, supongamos que un grupo de médicos pone en marcha un ensayo para evaluar la eficacia de un nuevo fármaco para la esclerosis múltiple, en comparación con el tratamiento estándar (interferón β). Si estos médicos, que con probabilidad tenderán a ser unos entusiastas de la posible eficacia del nuevo tratamiento, participan en la asignación de los pacientes, podrían influir en el proceso favoreciendo que sus propios pacientes o los pacientes de menor gravedad fuesen asignados al nuevo tratamiento. Por otro lado, las pruebas de significación estadística que se utilizan en los ensayos están basadas precisamente en la asunción de que el método de asignación ha sido aleatorio. Finalmente, frente a otros métodos de control de factores de confusión utilizados en las fases de diseño (muestreo estratificado, emparejamiento) o análisis (ajuste de tasas, regresión múltiple), la aleatorización presenta la gran ventaja de que controla tanto factores conocidos como desconocidos.

La eficacia de la aleatorización depende del tamaño muestral del estudio. Cuando éste no es lo suficientemente grande, —a modo de orientación estableceremos el punto de corte en 100 sujetos aunque esto también depende del número de grupos analizados—, no se puede garantizar el equilibrio en el tamaño de los grupos ni la distribución homogénea de los factores de confusión. Por ello, a la hora de presentar los resultados de un ensayo el primer paso debe consistir en mostrar una tabla con la distribución de los factores pronósticos conocidos en los grupos de intervención y control. Una distribución homogénea del número de sujetos y de los factores conocidos en ambos grupos constituye una prueba de que la aleatorización ha sido eficaz y nos indica que, con toda probabilidad, los grupos también serán comparables respecto de posibles factores pronósticos que por ser desconocidos no se hayan podido medir, permitiendo atribuir con seguridad el efecto observado a la intervención bajo estudio.

Un buen método de aleatorización debe de cumplir los siguientes requisitos para garantizar la calidad del mismo y la eficacia del proceso:

1. La *asignación* tiene que ser *ciega,* es decir, desconocida por el investigador hasta el momento en que el sujeto da su autorización para participar en el estudio. Desde el punto de vista operativo, es recomendable utilizar sobres opacos y sellados para evitar un posible sesgo en caso de que se descubra antes de tiempo la intervención a la que va a ser asignado el próximo sujeto que entre a formar parte del estudio.
2. La secuencia de asignación debe resultar impredecible, es decir, el investigador debería ser incapaz de adivinar con antelación la intervención a la que va a ser asignado el próximo sujeto. En ciertas ocasiones, adivinar la secuencia de aleatorización se convierte en un reto para la investigador.
3. El proceso debe estar basado en propiedades matemáticas conocidas y ser reproducible. Los métodos de aleatorización sistemáticos (asignación según el orden de incorporación al estudio, la fecha de nacimiento, etc.) no son válidos ya que, aunque son aleatorios, también son predecibles y, además, no están basados en un modelo matemático ni son reproducibles.

2.1. Métodos de aleatorización

Los métodos de aleatorización se dividen en dos grandes grupos. Por un lado, los métodos de asignación fija, en los que la probabilidad de ser asignado a uno u otro grupo no varía a lo largo del el período de inclusión de pacientes en el ensayo. Por otro lado, los métodos de asignación adaptativa, en los que esa probabilidad varia con el tiempo en función de la existencia de desequilibrios en el número de sujetos o en la distribución de los factores pronósticos en los grupos comparados o de la respuesta producida ante la intervención en los sujetos previamente incluidos en el ensayo.

Dentro de los métodos de asignación fija, la *aleatorización simple* constituye el procedimiento más elemental, intuitivo y fácil de aplicar. Para ilustrar este método nos referiremos al ejemplo mencionado previamente sobre la investigación de una nueva terapia para la esclerosis múltiple. Un método de aleatorización simple consistiría en lanzar una moneda al aire para cada nuevo paciente incluido en el ensayo y asignarle el nuevo tratamiento cuando apareciese cara y el interferón β cuando fuera cruz. El mismo método puede llevarse a cabo de forma menos rudimentaria con una tabla de números aleatorios o, con mayor nivel de sofisticación, mediante la generación de números aleatorios por medio de un programa informático, técnica muy conveniente cuando el tamaño de la muestra es elevado, en estudios multicéntricos y cuanto mayor es el número de intervenciones a aplicar.

La ventaja de la aleatorización simple es su sencillez. La desventaja, que ya hemos señalado previamente, es que si el número de sujetos del ensayo no es lo suficientemente grande, no se puede garantizar el equilibrio en el tamaño de los grupos ni la distribución homogénea de los factores de confusión. El distinto tamaño de los grupos puede afectar negativamente a las pruebas de significación estadística. Para evitar este inconveniente puede utilizarse la *aleatorización por bloques,* que consiste en definir bloques de sujetos de un tamaño predeterminado en los que por diseño se fuerza a que el número de sujetos asignados a cada una de las intervenciones bajo estudio sea el mismo. Por ejemplo, si en el estudio de la esclerosis múltiple hacemos bloques de 8 pacientes, en cada bloque habrá necesariamente 4 pacientes asignados al nuevo tratamiento y otros 4 al interferón β. Supongamos que al lanzar la moneda por primera vez obtenemos una cara: el primer sujeto de ese bloque es asignado al nuevo tratamiento. Continuamos lanzando la moneda y por efectos del azar obtenemos 4 cruces seguidas: los 4 siguientes pacientes del bloque son asignados al interferón. A partir de ese momento ya no volveríamos a lanzar la moneda, sino que asignaríamos "por decreto" a los 3 pacientes restantes para completar ese bloque al nuevo tratamiento. El problema con este método es que el investigador podría adivinar con antelación el tratamiento al que van a ser asignados los 3 últimos pacientes del bloque, lo que podría provocar un sesgo de asignación si el investigador evitase incluir en ese momento a los pacientes más graves favoreciendo que los pacientes menos graves fuesen asignados al nuevo tratamiento cuya eficacia se quiere demostrar. Una forma de evitar este problema consiste en hacer bloques de distinto tamaño cuyo orden también se asigna de forma aleatoria.

El problema derivado del posible desequilibrio en los factores pronósticos puede ser abordado a través de técnicas analíticas de ajuste o mediante la *aleatorización estratificada.* Para aplicar esta técnica es preciso medir previamente los factores pronósticos por los que se desea estratificar. El número de estratos es el producto del número de categorías de cada uno de los factores considerados; por ejemplo, si las variables a considerar son sexo (hombre-mujer) y estado civil (soltero-casado-separado-viudo), el número total de estratos es igual a ocho. Cuando un nuevo sujeto es incorporado al ensayo se le incluye en el estrato que le corresponda en función del sexo y estado civil (ej.: hombre soltero) y dentro de cada estrato se lleva a cabo un proceso de aleatorización independiente. La técnica de aleatorización dentro

de cada estrato puede ser simple o, más comúnmente, para evitar desequilibrios en el tamaño de los grupos de intervención dentro de cada estrato, se realiza una aleatorización por bloques.

3. ENSAYOS DE PREVENCIÓN PRIMARIA

Los ensayos de prevención primaria, como su propio nombre indica, tienen la finalidad de evaluar la eficacia de intervenciones preventivas aplicadas en sujetos sanos. La complejidad logística de este tipo de ensayos es enorme, ya que al no padecer ninguna enfermedad los sujetos no buscan atención sanitaria y, en consecuencia, es preciso visitarlos a nivel de campo (trabajo, escuela) o establecer centros desde los que llevar a cabo el estudio y a los que los participantes puedan acudir. Por otro lado, el riesgo de que se produzca el efecto de interés es muy bajo, por lo que requieren un tamaño muestral muy elevado para conseguir que ése desenlace se desarrolle en un número de casos suficiente. Ambas características explican por sí solas el alto coste de este tipo de estudios, que se dividen en ensayos de campo y ensayos de intervención comunitaria.

3.1. Ensayos de campo

Los ensayos de campo son estudios experimentales de prevención primaria en los que la intervención se aplica a nivel individual; es decir, cada sujeto que entra en el ensayo es randomizado de forma individual a una de las intervenciones estudiadas. Los elevados costes de los ensayos de campo limitan su uso al estudio de factores preventivos de enfermedades bien muy frecuentes, bien extremadamente graves. En general se centran en intervenciones dietéticas para prevenir deficiencias de algún nutriente (flúor, yodo, calcio) o dirigidas a modificar los estilos de vida (actividad física, tabaco, dieta, obesidad) relacionados con enfermedades crónicas muy prevalentes (caries, bocio, osteoporosis, cardiovasculares).

El mayor experimento llevado a cabo hasta ahora en humanos fue un ensayo de campo realizado en 1954, en el que se administró vacuna Salk o placebo a 800.000 escolares estadounidenses. A pesar de la baja incidencia de la poliomielitis (1/2.000 anual), la gravedad de la misma justificó los esfuerzos y los recursos que fue necesario poner encima de la mesa para realizar el ensayo. Los resultados fueron concluyentes: la incidencia de poliomielitis y de enfermedad paralítica en el grupo de intervención se redujeron un 50% y las 4 muertes por poliomielitis se produjeron en el grupo control.

3.2. Ensayos de intervención comunitaria

El ensayo comunitario de intervención es una extensión del ensayo de campo, con la particularidad de que la intervención se asigna a grupos de personas en lugar de hacerlo de forma individual. El nivel de agregación de los grupos depende de la factibilidad de exponer todos sus miembros a la intervención de forma simultánea, como la familia para intervenciones dietéticas, la escuela para programas educativos o la empresa para intervenciones de carácter medioambiental. En otros casos, la pertenencia a un determinado grupo, como hospitales, unidades del ejército o comunidades religiosas, puede favorecer un mayor nivel de cumplimiento de la intervención en estudio.

Este tipo de ensayos permitieron demostrar la eficacia del suministro de suplementos de flúor a través de la red de agua de consumo en la prevención de caries dental. En los años 30, diversos estudios ecológicos en los Estados Unidos encontraron una asociación inversa entre la concentración de flúor en el agua de abastecimiento de la comunidad y el índice CAOD en niños de 12 a 14 años. Una vez que se determinó que 1 ppm. era la concentración de flúor

óptima y segura, se realizó un ensayo comunitario (sin grupo control) en tres comunidades en 1945, observándose una reducción del 50-60% de la prevalencia de caries en las tres comunidades. Estudios posteriores con grupo control confirmaron los resultados observados.

La eficacia de la aleatorización en este tipo de ensayos depende del tamaño del grupo de aleatorización en relación con el tamaño total del estudio: cuanto menor sea esta fracción más probable será que la aleatorización de lugar a que los factores pronósticos se distribuyan de forma similar en los grupos de intervención. El mejor escenario posible se produce cuando hay tantos grupos como individuos (ensayo de campo) y el peor cuando sólo hay dos grupos, en cuyo caso da lo mismo aleatorizar o no.

4. ENSAYO CLINICO

Un ensayo clínico es un experimento planificado en el que, de forma prospectiva, se comparan dos o más intervenciones (o intervención y placebo) preventivas, curativas o rehabilitadoras, que son asignadas de forma individualizada y aleatoria a un grupo de pacientes para determinar su eficacia. Además, tanto la selección de los sujetos como los períodos de tratamiento y seguimiento han de tener lugar simultáneamente en todos los grupos. Dado que los participantes en un ensayo clínico son, por definición, sujetos enfermos (pacientes), las intervenciones preventivas a las que nos referimos son secundarias, dirigidas a evitar las secuelas de la enfermedad, incluyendo la muerte. Por ejemplo, una dieta rica en fibra para prevenir la recurrencia de pólipos adenomatosos de colón o una dieta rica en ácido α-linolénico para reducir la tasa de mortalidad post-infarto de miocardio.

En 1935 Sir Austin Bradford Hill formuló los principios de los ensayos clínicos controlados y aleatorizados en medicina, principios que se mantienen totalmente vigentes hoy en día. El primer ensayo clínico con un grupo control randomizado de forma apropiada se llevó a cabo en 1948 para evaluar la eficacia de la estreptomicina en el tratamiento de la tuberculosis pulmonar. Desde entonces, los ensayos clínicos han proporcionado pruebas empíricas de la eficacia de muchas intervenciones médicas, contribuyendo enormemente a la mejora de la prescripción y la utilización de medicamentos y otro tipo de intervenciones médicas.

Los ensayos clínicos evalúan las intervenciones en condiciones muy diferentes a las de su aplicación habitual, ya que los pacientes están sometidos a un estrecho seguimiento y vigilancia médica y no suelen ser representativos de todo el espectro de posibles sujetos que padecen la enfermedad en cuestión. Por lo tanto, un ensayo está evaluando el efecto de la intervención en condiciones ideales (eficacia). Sin embargo, no puede evaluar como se comportará esa misma intervención en condiciones reales de uso (efectividad) ni su relación coste-beneficio (eficiencia), aspectos que se contemplan en el marco de la investigación de resultados. Los pacientes seleccionados para realizar un ensayo representan habitualmente al grupo más accesible al equipo investigador, con menos patologías asociadas y con mayores probabilidades de cumplir con el protocolo del estudio. Además, los pacientes que aceptan participar son en general sujetos motivados y más concienciados con su estado de salud, por lo que es probable que el grado de adhesión a una terapia concreta o a una modificación dietética sea mayor del que observaríamos en el resto de pacientes. Por otro lado, la estrecha vigilancia a la que están sometidos estos pacientes mientras dura el estudio hace que los incumplimientos o abandonos del tratamiento sean menos probables que en condiciones habituales.

Al principio de este apartado hemos señalando que un ensayo clínico nunca se debe improvisar, sino que debe surgir siempre a partir de un proceso de planificación que se plasme en un protocolo escrito, que es el que determina las acciones a seguir. En la tabla 10-2 se señalan los apartados que de forma general debería contemplar el protocolo de cualquier ensayo clínico. Este listado representa una guía orientativa, que puede ser adaptada a cada ensayo particular en función de las circunstancias del mismo. En el caso concreto de los ensayos clínicos con medicamentos, la legislación española ha establecido mediante Real Decreto el procedimiento a seguir y los apartados que se deben incluir en el protocolo del mismo (RD 223/2004, de 6 de febrero, por el que se regulan los ensayos clínicos con medicamentos). En las siguientes secciones vamos a profundizar sobre algunos de los aspectos más importantes de la planificación y desarrollo de un ensayo clínico.

Tabla 10-2. Protocolo de un ensayo clínico

1. Antecedentes y estado actual del tema
2. Objetivos generales y específicos
3. Diseño del ensayo
4. Criterios de selección de los pacientes
5. Reclutamiento y aleatorización de los pacientes
6. Descripción de la intervención
7. Métodos de evaluación de la respuesta
8. Tamaño de la muestra requerido
9. Monitorización del ensayo
10. Impresos de recogida y manejo de datos: manual de operaciones
11. Desviaciones del protocolo
12. Análisis estadístico previsto
13. Aspectos éticos: consentimiento informado
14. Aspectos administrativos

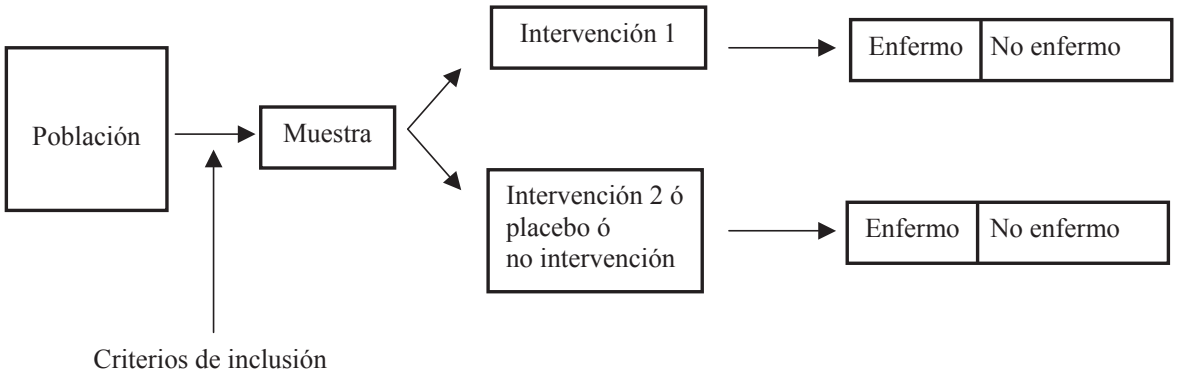
4.1. Diseño

4.1.1. Ensayo clínico de grupos paralelos

El ensayo de grupos paralelos constituye, por su sencillez, el diseño básico para realizar un ensayo clínico. En este modelo, cada paciente es asignado al mismo tiempo y de forma aleatoria a una intervención diferente y tras un período de seguimiento variable, se procede a medir la respuesta (Figura 10-1). En general, se incluyen dos grupos de comparación, dada la dificultad de reclutar suficiente número de pacientes por grupo cuando se aumenta el número de tratamientos a comparar. A veces, este diseño puede contemplar un período de preinclusión previo a la aleatorización, que tiene por objeto disminuir las tasas de no cumplimiento de la intervención o de abandono del estudio y, en ocasiones, estabilizar los

valores de las variables a través de las cuales se va a medir la respuesta. Durante el período de preinclusión, todos los individuos son sometidos a la misma intervención, habitualmente la del grupo control, excluyendo del estudio a los sujetos no cumplidores. Mensink y cols. realizaron un ensayo clínico de grupos paralelos para comparar el efecto de los ácidos grasos monoinsaturados y poliinsaturados sobre el perfil lipídico plasmático. Para ello, establecieron un periodo de preinclusión durante el que los sujetos fueron asignados a una dieta rica en ácidos grasos saturados, permitiendo valorar el grado de cumplimiento de la dieta y estabilizar los lípidos sanguíneos. Una vez finalizado este periodo los sujetos cumplidores fueron asignados de forma aleatoria a una dieta rica en ácidos grasos mono o poliinsaturados respectivamente.

Figura 10-1. Ensayo clínico de grupos paralelos

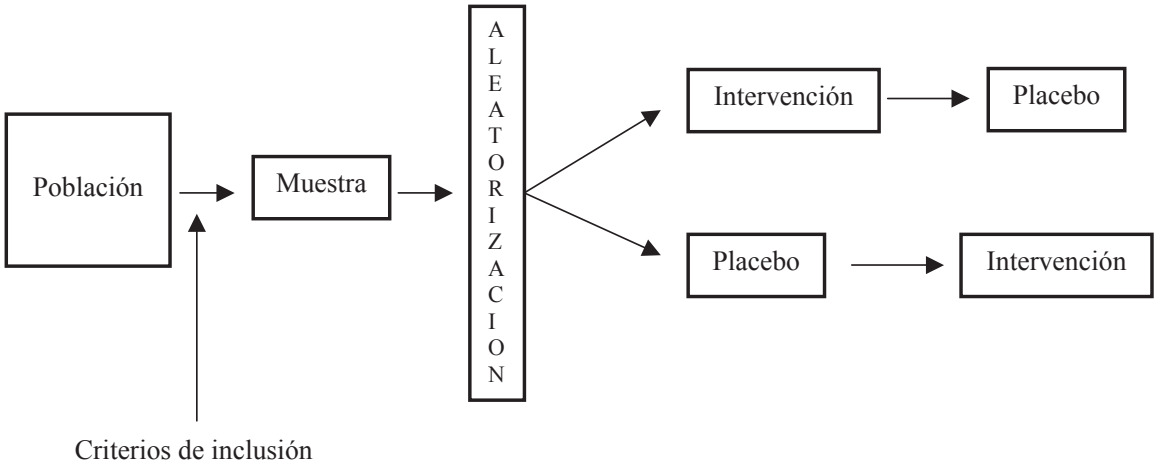


4.1.2. Ensayo clínico cruzado

El problema principal de los ensayos de grupos paralelos es que habitualmente los pacientes presentan una gran variación en el estado inicial de la enfermedad, las variables pronósticas y en su respuesta a la intervención. Esto implica que se necesita un número importante de pacientes en cada brazo del ensayo para poder estimar de forma fiable la magnitud de la diferencia de efecto de las intervenciones estudiadas. Una forma de abordar este problema consiste en diseñar un ensayo cruzado, en el que cada paciente es sometido a todas las intervenciones durante periodos de tiempo sucesivos, realizándose después una comparación de tratamientos intra-paciente. Estas comparaciones son más eficientes, ya que se reduce la variablilidad intra-grupo al servir cada individuo como su propio control, por lo que se obtienen resultados más precisos con tamaños muestrales reducidos. Como contrapunto, una muestra pequeña puede condicionar la eficacia de la aleatorización para distribuir las variables de interés de forma homogénea en los grupos a comparar.

En este diseño cada individuo es asignado aleatoriamente a una secuencia temporal de dos o más tratamientos, al objeto de controlar el posible efecto período derivado de posibles cambios en otras variables predictoras a lo largo del tiempo (figura 10-2). Es decir, lo que se aleatoriza no es el grupo de intervención al que va a ser asignado el paciente, sino el orden en el que va a recibir las intervenciones del estudio.

Figura 10-2. Ensayo clínico cruzado



La principal limitación de este diseño es que las condiciones de aplicación del mismo son muy restrictivas: únicamente está indicado en enfermedades estables con tratamientos no curativos de respuesta rápida y transitoria, requisitos que se dan en algunos procesos crónicos y recidivantes como algunas de las enfermedades reumáticas y alérgicas. Además, asume ausencia de efectos residuales de la medicación ("***carry-over***"), por lo que habitualmente es necesario intercalar fases de lavado (sin medicación activa) entre los diferentes periodos de tratamiento para poder cumplir este requisito. Las variables resultado suelen tener carácter intermedio (no incluyen la curación o la muerte), por lo que este diseño es más útil en evaluaciones iniciales de la eficacia de los tratamientos -ensayos fase II-, siendo de elección en el estudio de bioequivalencia de diferentes formulaciones o dosis del mismo fármaco en voluntarios sanos. Por último, otra limitación destacable de este diseño se deriva de la complejidad de su análisis estadístico (figura 10-3).

4.1.3. Ensayo clínico factorial

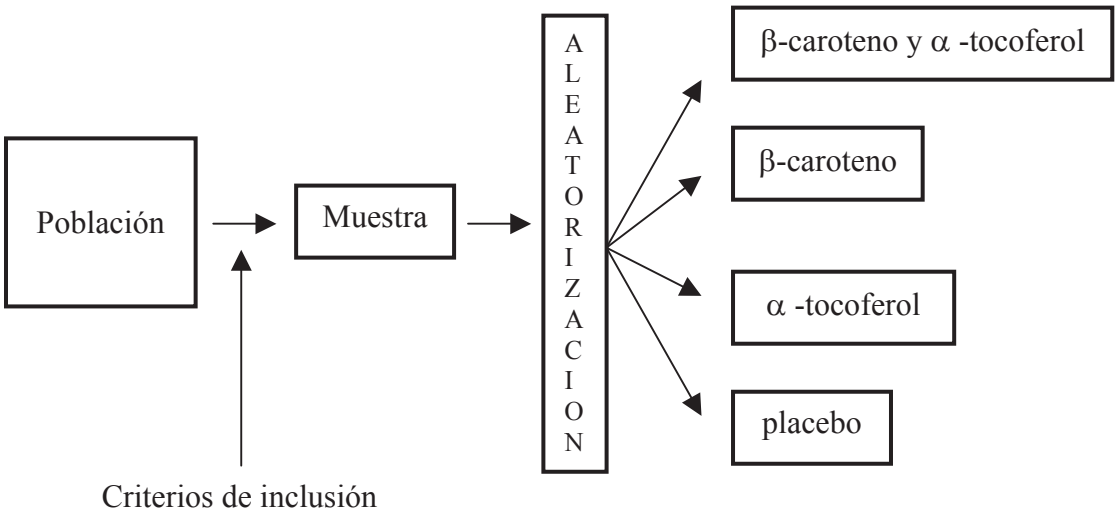
Cuando se desea evaluar varias intervenciones en un mismo estudio, una alternativa consiste en diseñar un ensayo de grupos paralelos con un brazo diferente del ensayo para cada una de las intervenciones y otro brazo para el placebo. Por ejemplo, un ensayo de prevención cardiovascular que pretendiese evaluar de forma conjunta una intervención dietética (sustitución de alimentos ricos en AGS por aceite de oliva), un programa de actividad física y una intervención educativa sobre técnicas psicológicas para afrontar el estrés. Una forma de reducir de forma drástica el elevado tamaño muestral que se necesitaría en este estudio consiste en diseñar un ensayo paralelo de dos grupos y aplicar todas las intervenciones en el mismo brazo del ensayo, dejando el otro brazo para el placebo. Esta alternativa tiene la ventaja adicional de que al aplicar varias intervenciones a la vez , habitualmente aumentará la magnitud del efecto global, con el consiguiente incremento de la potencia del estudio y disminución del tamaño muestral requerido. Obviamente, la desventaja de este diseño es que no podemos evaluar los efectos de las diferentes intervenciones por separado. Además, en el caso poco probable de que alguna intervención tuviese el efecto contrario al esperado, podríamos encontrarnos con un efecto global nulo como resultado de la suma de los efectos contrapuestos de diferentes intervenciones.

La forma de abordar la necesidad de reducir el tamaño muestral cuando se desea evaluar varios tratamientos (con mecanismos de acción y efectos independientes) a la vez, consiste en realizar un ensayo clínico factorial. En este diseño cada paciente es asignado aleatoriamente a una combinación de dos o más tratamientos o a una combinación de diferentes dosis de un

grupo de tratamientos. De esta forma se incrementa de forma extraordinaria la eficiencia del estudio, ya que permite evaluar varios tratamientos con el mismo número de individuos que hubieran sido necesarios para evaluar uno sólo. Por ejemplo, un grupo de investigación finlandés diseñó un ensayo factorial para evaluar de forma simultánea la eficacia de los suplementos dietéticos de β-caroteno y α-tocoferol en la prevención del cáncer de pulmón. Si se hubiese utilizado un diseño paralelo con tres brazos, cada intervención hubiese sido aplicada en un tercio de los individuos, mientras que con el diseño factorial tanto el β-caroteno como el α-tocoferol fueron suministrados a la mitad de los sujetos del estudio.

El diseño factorial sirve también para evaluar posibles interacciones entre los efectos de las intervenciones estudiadas. Cuando se utiliza con este objeto este diseño pierde su atractivo en términos de eficiencia, ya que la interpretación de los resultados se complica y son precisos tamaños muestrales más elevados para evaluar dicha interacción.

Figura 10-3. Ensayo clínico factorial



4.2. Selección de los participantes

La selección de los sujetos que participarán en el estudio debe llevarse a cabo de forma cuidadosa. Para ello, hay que empezar por definir de forma precisa el perfil de los sujetos susceptibles de participar en el estudio, por ej.: sujetos de ambos sexos, de 25 a 40 años de edad, con sobrepeso (IMC entre 25 y 30) y residentes en Madrid. A continuación se deben definir los criterios de exclusión, en función de aspectos que pudieran comprometer la participación en el estudio (ej.: enfermedades incapacitantes física o mentalmente) o que pudieran interaccionar con la intervención bajo estudio y afectar la variable de interés. En cualquier caso, hemos de intentar ser lo menos restrictivos posible a la hora de establecer los criterios de inclusión y exclusión, al objeto de evitar que tengamos problemas para reclutar un número apropiado de sujetos. Un ensayo clínico con un número pequeño de pacientes conlleva un riesgo considerable de fracasar en el intento de demostrar una diferencia entre las intervenciones que se están comparando cuando ésta realmente existe. Un aspecto adicional a tener en cuenta al definir el perfil de los participantes es el de la representatividad de la muestra. Aunque a efectos de comparabilidad entre los grupos de intervención la representatividad no es relevante, cuando diseñamos un ensayo clínico no sólo deseamos demostrar la eficacia de la intervención en los pacientes del ensayo, sino en todos aquellos sujetos o pacientes susceptibles de recibir la intervención en el futuro. Por ello, dentro de las posibilidades del equipo investigador, y aún a costa de reducir la extensión o calidad de las observaciones realizadas a los participantes, es

necesario intentar que la muestra sea lo más representativa posible de aquélla población a la que se pretende aplicar la intervención.

4.3. Enmascaramiento

El enmascaramiento, o carácter ciego de un ensayo clínico, es el conjunto de medidas o precauciones que se toman con el fin de que los que intervienen en el mismo (paciente, médico, evaluador, ..) desconozcan el tratamiento al que el paciente ha sido asignado, ya que este conocimiento podría introducir sesgos en el desarrollo, análisis y evaluación de un ensayo clínico. La presentación de las intervenciones enmascaradas deberá ser idéntica, diferenciándose únicamente por la sustancia activa que contienen. Hay que tener en cuenta varios aspectos que tienen que ser idénticos en ambas intervenciones: vía y pauta de administración , forma, tamaño, color, sabor, textura, etc..

El enmascaramiento del responsable de evaluar la respuesta (paciente, médico o un tercero) a la intervención es particularmente importante cuando la variable resultado es subjetiva (ej: valoración de la intensidad del dolor, análisis de una radiografía, ..), ya que el conocimiento de la intervención asignada puede afectar la percepción sobre la variable respuesta. Los resultados más objetivos, como la muerte, son relativamente insensibles a este tipo de sesgo. Por ej: si uno de los médicos participantes en el estudio, es un entusiasta de la eficacia del nuevo tratamiento bajo estudio, podría transmitir este entusiasmo a sus pacientes, y esto sin duda podría alterar la valoración de la variable respuesta, tanto por parte del médico como por parte del paciente.

Lamentablemente, no siempre es posible aplicar el enmascaramiento, por ej. en el caso de que la intervención sea una dieta o una intervención quirúrgica. Por otro lado, en ocasiones un tratamiento puede tener efectos secundarios bien conocidos que pueden permitir al paciente realizar la identificación del mismo. Por ej.: la rifampicina tiñe de rojo la orina. Además, nunca deben enmascararse los tratamientos si por ello sometemos a algún tipo de riesgo a los pacientes. Por todo ello, para poder hacer una valoración lo más objetiva y realista posible de los resultados del ensayo clínico en relación con la eficacia de la intervención, el investigador debe describir no solo el método de enmascaramiento y las violaciones del mismo, sino también los aspectos comentados previamente, en relación con la imposibilidad de aplicarlo o los posibles riesgos de que no sea eficaz.

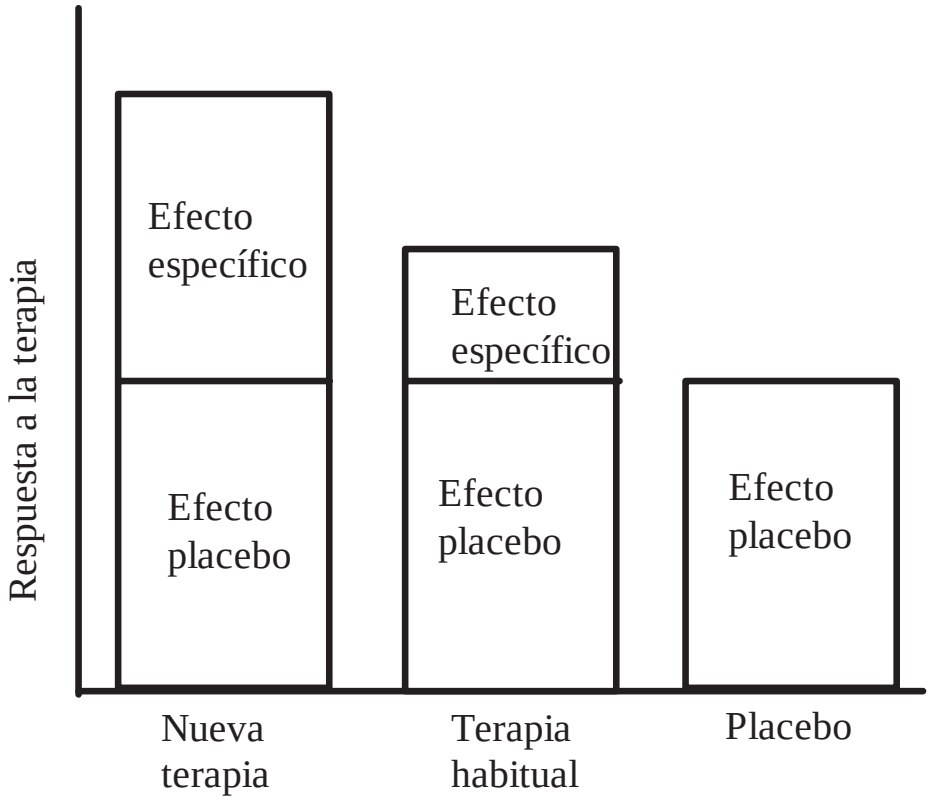
En función del grado de enmascaramiento decimos que un ensayo clínico es abierto cuando el paciente y el investigador conocen la intervención asignada, simple ciego, cuando el paciente es el único que desconoce la intervención asignada y doble ciego, cuando el paciente, el responsable de su cuidado (equipo de tratamiento) y el evaluador de la variable respuesta (investigador) desconocen el grupo de tratamiento asignado. Aunque en el doble ciego hay en realidad tres agentes implicados (paciente, equipo de tratamiento y evaluador), se mantiene ésa denominación porque en muchas ocasiones el médico que trata al paciente es asimismo el responsable de la evaluación de la variable respuesta.

4.4. El placebo

La administración de una intervención produce un efecto terapéutico sobre el individuo, denominado efecto placebo, independientemente del principio activo que contenga la misma. Es decir, incluso la administración de un principio inactivo inocuo produce un efecto medible. El simple hecho de ser observado, de participar en un estudio, produce una modificación positiva en el estado clínico de los pacientes, denominada “efecto Hawthorne”. A esto hay que añadir el poder de sugestión derivado del hecho de recibir una terapia, aunque ésta sea farmacológicamente inactiva, y el de la presencia de un médico, o investigador, que aplica la misma.

El efecto global de cualquier intervención tiene dos componentes: el efecto placebo y el efecto específico de la misma (Figura 10-4). Por ello, al estudiar la eficacia de una nueva terapia, es preciso aplicar una intervención alternativa al grupo control, que permita controlar el efecto placebo y discernir el efecto específico de la nueva terapia. Por razones éticas, cuando exista una terapia eficaz disponible para la condición bajo estudio no aplicaremos un placebo a los pacientes del grupo control, sino ese tratamiento habitual. La razón principal para utilizar el placebo consiste en conseguir que la actitud de los individuos ante la investigación en los grupos de intervención y control sea la misma. Asimismo, proporciona la oportunidad de realizar ensayos doble ciegos y evitar de esa forma que se produzcan sesgos en el desarrollo, análisis y evaluación del ensayo, tal como hemos señalado en el apartado anterior.

Figura 10-4. El efecto placebo



4.5. Evaluación de la respuesta y presentación de los resultados

El protocolo de un ensayo clínico debe distinguir claramente entre la variable principal de respuesta, que va a determinar el tamaño muestral y la posible eficacia de la intervención, y las variables de valoración secundarias. En un ensayo sobre los efectos para la salud de una dieta vegetariana, las variables de mayor interés serían aquellas que midieran el efecto sobre la mortalidad general o la debida a enfermedades crónicas como cáncer o cardiovasculares. Sin embargo, para poder detectar un número aceptable de defunciones sería necesario seleccionar una muestra muy numerosa y de alto riesgo y mantener el ensayo por un largo período de tiempo, condiciones que pueden limitar la factibilidad del estudio. A más corto plazo, pueden utilizarse otras variables de valoración principal como la tensión arterial, el índice de masa corporal o el nivel de oxidación de determinadas sustancias.

En general, el análisis principal de un ensayo clínico debe basarse en el principio de intención de tratar, que incluye a cada paciente en su grupo de aleatorización independientemente de los cambios que pudieran ocurrir a posteriori: incumplimiento de la intervención, cambio de tratamiento, etc.. Aunque este tipo de análisis tiene una mayor tendencia a obtener resultados nulos, aumentando las dificultades para demostrar la eficacia de cualquier nueva intervención, existen dos poderosas razones para llevarlo a cabo. Por un lado, con este tipo de análisis se pretende preservar lo conseguido con la aleatorización, que los grupos sean comparables, lo cual constituye el fundamento para poder obtener conclusiones válidas (no sesgadas) sobre la eficacia de una intervención. Por otro lado, se trata de evaluar el efecto de la intervención en condiciones lo más parecidas posibles a las de la vida real (efectividad) y no sólo teniendo en cuenta las condiciones ideales (eficacia) que proporcionan el subgrupo de pacientes que han seguido el protocolo de forma estricta.

Tradicionalmente, la medición del efecto de una intervención podía llevarse a cabo mediante medidas absolutas, como la reducción del riesgo atribuible, o relativas, como la reducción porcentual del riesgo relativo. De forma más reciente han surgido otras medidas de carácter complementario, que ponen el énfasis en facilitar la tarea del clínico a la hora de orientar el tratamiento de sus pacientes. Estas medidas de carácter más pragmático incluyen el porcentaje de pacientes libres de la condición bajo estudio y el número necesario de pacientes a tratar (NNT). El NNT constituye el número de pacientes que tendríamos que tratar para evitar que se produjera un caso con la enfermedad o condición estudiada y se calcula como el inverso de la reducción del riesgo atribuible. La presentación de los resultados mediante un único indicador, como la reducción porcentual del riesgo relativo, puede producir una impresión magnificada sobre el efecto del tratamiento, por lo que actualmente se recomienda que al presentar los resultados de un ensayo clínico se utilice la reducción porcentual del riesgo relativo acompañada de alguna magnitud que mida la reducción del riesgo en términos absolutos, particularmente el NNT. Para ilustrar esta última afirmación, nos basaremos en el ejemplo descrito por Cook y Sacket, a partir de una revisión de ensayos clínicos sobre tratamientos farmacológicos para la hipertensión arterial realizada por Collins y cols.. Los estudios fueron divididos en dos grupos mutuamente exclusivos: aquéllos en los que todos los pacientes tenían una presión diastólica al inicio del ensayo menor de 110 mmHg (hipertensión ligera) y aquéllos en los que todos los pacientes tenían una presión diastólica al inicio del estudio menor de 115 mmHg (hipertensión moderada). Como puede apreciarse en la tabla 10-3, la reducción porcentual del riesgo relativo de padecer un ACV durante los 5 años siguientes con tratamiento farmacológico fue similar en los dos grupos, del 40%, lo que podría inducir a pensar que la estrategia terapéutica debería ser igual en ambos grupos. Sin embargo, la reducción absoluta del riesgo fue mucho menor en los sujetos con hipertensión ligera, dado que este grupo de pacientes tenía una tasa de ACV en ausencia de tratamiento mucho más baja que el grupo con hipertensión moderada. Traducido en términos de NNT, mientras en un caso se necesitaba tratar 13 pacientes para evitar un ACV, en el otro era preciso tratar a 167 pacientes para obtener el mismo resultado. En consecuencia, parece que la estrategia terapéutica debería ser diferente en esos dos grupos de pacientes.

Tabla 10-3. Cálculo de la reducción del riesgo y del NNT para evitar un caso de ACV en pacientes con hipertensión

	ACV en 5 años		Reducción del RR	Reducción del RA	NNT
Hipertensión	Grupo control	Grupo tratamiento	(Pc – Pt) / Pt	Pc – Pt	1 / (Pc – Pt)
Moderada					
Tasa de ACV (P)	0,20	0,12	0,40	0,08	13
N	16.778	16.898			
Ligera					
Tasa de ACV (P)	0,015	0,009	0,40	0,006	167
N	15.165	15.238			

* Tomado de BMJ 1995; 310: 452-454

Por último, la importancia de una adecuada presentación de los resultados de los ensayos clínicos ha llevado a la elaboración de la "Declaración CONSORT", un informe con recomendaciones para mejorar la calidad de la presentación de los resultados de ensayos clínicos elaborado por investigadores, estadísticos, epidemiólogos y editores biomédicos y apoyado por numerosas revistas biomédicas y grupos editoriales. La declaración recoge los aspectos (22 ítems) que han de ser contemplados y la forma de presentarlos en los diferentes apartados (título y resumen, introducción, métodos, resultados y discusión) del informe sobre un ensayo clínico de dos grupos paralelos (tabla 10-4). La declaración, junto con las extensiones de la misma para los diferentes tipos de ensayos clínicos (en función de datos, intervenciones o diseños alternativos) se pueden consultar en la siguiente dirección: http://www.consort-statement.org/consort-statement.

Tabla 10-4. Declaración CONSORT: Lista de ítems a incluir en el informe sobre un ensayo clínico

Apartado	Ítem	Descripción
Titulo y resumen	1	Método utilizado para asignar los pacientes a las intervenciones (ej.: "asignación aleatoria", "aleatorizado" o "asignado aleatorizadamente").
Introducción		
Fundamento	2	Fundamento científico y explicación del objetivo principal.
Métodos		
Participantes	3	Criterios de inclusión-exclusión. Área de referencia y lugar donde se recogerán los datos.
Intervenciones	4	Información precisa y detallada de la intervención de cada grupo y de cómo y cuando fueron realmente administradas.
Objetivos	5	Hipótesis y objetivos específicos
Respuesta	6	Definir claramente las variables de respuesta principales y secundarias y, en su caso, cualquier método utilizado para mejorar la calidad de las mediciones (ej.: observaciones múltiples, entrenamiento de los evaluadores, etc.).
Tamaño muestral	7	Método de determinación del tamaño muestral y, en su caso, descripción de análisis interinos y criterios para parar el ensayo.
Aleatorización		
- Generación de la secuencia	8	Método utilizado para generar la secuencia de asignación aleatoria, incluyendo detalles de cualquier restricción aplicada (ej.: estratificación, aleatorización por bloques).
- Enmascaramiento de la asignación	9	Método utilizado para aplicar la secuencia de asignación aleatoria (ej.: sobres numerados o central telefónica), aclarando si la secuencia estuvo enmascarada hasta que las intervenciones fueron asignadas.
-Aplicación	10	Personas encargadas de generar la secuencia de asignación, reclutar a los participantes y asignarlos al grupo correspondiente.
Enmascaramiento	11	Grado de enmascaramiento de los participantes y de los encargados de aplicar la intervención y evaluar la respuesta y forma de evaluar el éxito del mismo.
Métodos estadísticos	12	Métodos estadísticos utilizados para comparar las variables principales de respuesta en los grupos: métodos para análisis adicionales, como análisis de subgrupos y análisis ajustados.

Apartado	Ítem	Descripción
Resultados		
Flujo de participantes	13	Flujo de participantes a través de cada etapa (un diagrama es intensamente recomendado). Específicamente, declarar para cada grupo el número de participantes que han sido asignados aleatoriamente, que han recibido el tratamiento asignado, que han cumplido el protocolo y que han participado en el análisis de la variable principal. Describir las desviaciones del protocolo y sus razones.
Reclutamiento	14	Fechas de los períodos de reclutamiento y seguimiento.
Datos basales	15	Características basales demográficas y clínicas de cada grupo.
Número analizado	16	Número de participantes (denominador) de cada grupo incluidos en cada análisis. Declarar si el análisis fue por "intención de tratar". Proporcionar los resultados en números absolutos siempre que sea posible.
Respuesta y Estimación	17	Para cada variable de respuesta principal o secundaria, presentar un resumen de los resultados para cada grupo, el tamaño estimado del efecto y su precisión (ej.: 95% CI).
Análisis auxiliares	18	Abordar la multiplicidad declarando cualquier otro análisis realizado, incluyendo análisis de subgrupos y ajustados, e indicando si éstos eran predeterminados o exploratorios.
Efectos adversos	19	Todos los efectos adversos importantes o efectos secundarios en cada grupo de intervención.
Discusión		
Interpretación	20	Interpretación de los resultados, teniendo en cuenta la hipótesis del estudio, las fuentes de imprecisión o de sesgos potenciales y los peligros asociados con la multiplicidad de análisis y respuestas.
Generalizabilidad	21	Generalizabilidad (validez externa) de los hallazgos del ensayo.
Evidencia global	22	Interpretación general de los resultados en el contexto de la evidencia actual.
Traducido y adaptado de Lancet 2001; 357:1191-1194		

5. REFERENCIAS BIBLIOGRÁFICAS

BAKKE OM, CARNÉ X, GARCÍA ALONSO F. *Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica.* Barcelona: Doyma, 1994.

FRIEDMAN LM, FURBERG C, DEMETS DL. *Fundamentals of clinical trials.* 2ª ed. Littleton: PSG Publishing Company, 1985.

GALENDE I, SACRISTÁN JA, SOTO J. *Cómo mejorar la calidad de los ensayos clínicos*. Med Clin (Barc) 1983; 80: 768-771.

GUALLAR E, ROYO-BORDONADA MA. *Directrices del "Commitee propietary medicinal products": Estadística.* En Ergón ed. II Reunión de Actualización de ensayos clínicos con medicamentos. Madrid, 1996.

MARRUGAT J. *Los Estudios Experimentales.* En: Martinez Navarro F. Salud Pública, ed. McGraw-Hill Interamericana, 1997: 239-258.

MOHER D, SCHULZ K, ALTMAN D. *The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials.* Lancet 2001; 357: 1191-1194.

POCOCK SJ. *Clinical trials. A practical approach*. Chichester, GB: John Willey & Sons, 1983.

PORTA M, IBÁÑEZ LI, CARNÉ X, LAPORTE JR. *Principios del ensayo clínico*. Med Clin (Barc) 1983; 80: 768-771.

REAL DECRETO 223/2004, de 6 de abril, por el que se regulan los ensayos clínicos con medicamentos - BOE núm. 33, de 7 de febrero de 2004: 5429-5443.

XI. EPIDEMIOLOGÍA CLÍNICA.

Iñaki Imaz Iglesia
Jesús González Enríquez

1. INTRODUCCIÓN. CONCEPTO DE EPIDEMIOLOGÍA CLÍNICA Y APLICACIONES.

Los conceptos, herramientas y métodos de la epidemiología se han aplicado al estudio de la práctica clínica dando lugar a la *Epidemiología Clínica.* Esta disciplina pretende dar respuesta a las preguntas que surgen del contacto médico-paciente (observaciones, decisiones y evaluaciones clínicas) mediante un abordaje poblacional y no exclusivamente individual. El método clínico tradicional por el que un médico/a se plantea los problemas de su consulta considerando a cada paciente de forma individual y cada caso clínico igualmente relevante es superado mediante una metodología más generalizable.

La práctica clínica es el origen, fuente de preguntas y destino final de la Epidemiología Clínca. Se trata de aplicar el método científico a la resolución de problemas clínicos. La Epidemiología Clínica se ha definido de varias formas:

— Estudio de los determinantes y consecuencias de las decisiones médicas.
— Estudio de las decisiones diagnósticas, pronósticas y terapéuticas que se toman con respecto a los enfermos.
— Estudio de las variaciones en los desenlaces de la enfermedad y de las razones que conducen a ello.
— Ciencia y método de estudiar las decisiones óptimas en la medicina clínica, teniendo en cuenta las características epidemiológicas del paciente y su ambiente clínico externo, la patología que le concierne y los factores y maniobras a que se ve expuesto en su ambiente clínico, en especial las acciones médicas.

La Epidemiología Clínica surge como una disciplina autónoma de la mano de tres Universidades. En la Universidad de Yale de los EEUU con Alvan Feinstein, en la Universidad de Carolina del Norte (EEUU) con Fletcher y en la Universidad canadiense McMaster con Sackett.

Entre las múltiples áreas de la práctica clínica que aborda la Epidemiología Clínica (diagnóstico, tratamiento, pronóstico, prevención) en este manual se exponen la evaluación de pruebas diagnósticas y las decisiones clínicas. También se propone un análisis del caso del cribado poblacional de cáncer de mama mediante mamografía.

2. LA EVALUACIÓN DE LAS PRUEBAS DIAGNÓSTICAS.

Una buena prueba diagnóstica es aquella que es capaz de discriminar o clasificar correctamente a los miembros de la población en que se aplica, ofreciendo un resultado positivo en los enfermos o portadores de la condición clínica buscada, y negativo en los sanos o no portadores. La capacidad de la prueba de diagnosticar adecuadamente se evalúa fundamentalmente estudiando su validez y fiabilidad.

2.1. Validez de una prueba diagnóstica.

La validez es el grado en que una prueba mide lo que se supone debe medir y corresponde a la exactitud diagnóstica. Es la capacidad de la prueba de discriminar entre los casos y los no

casos para una enfermedad o condición clínica. La validación de una prueba diagnóstica se realiza mediante la comparación de sus resultados con los obtenidos mediante el mejor instrumento de medida del fenómeno estudiado, o con el verdadero resultado si éste es conocido, o sea mediante la comparación del test con un "estándar de oro".

Algunas veces el estándar de oro es una prueba diagnóstica relativamente simple y económica, como por ejemplo el cultivo de garganta para estreptococo ß–hemolítico del grupo A para validar la impresión clínica de faringitis estreptocócica. Sin embargo, más frecuentemente debe recurrirse a exámenes complejos, costosos o arriesgados (biopsia, radiología con contraste, autopsia).

En la tabla 11-1 se exponen los principales indicadores para evaluar la validez de una prueba diagnóstica. La *sensibilidad* de una prueba es la probabilidad de que un sujeto con la enfermedad sea adecuadamente clasificado. Una prueba será tanto más sensible cuanto menos falsos negativos (FN) proporcione. En las pruebas de cribado generalmente se utilizan test de alta sensibilidad para captar a casi todos los enfermos. También en enfermedades graves tratables, en las que no debe permitirse la pérdida de ningún caso.

Tabla 11-1. Relación entre una prueba diagnóstica y un estándar de oro.

		Enfermedad		
		+	-	
Prueba	+	a VP	b FP	**a+b**
	-	c FN	d VN	**c+d**
		a+c	**b+d**	**a+b+c+d**

$$Sensibilidad = \frac{a}{a+c} \qquad Especifidad = \frac{d}{b+d}$$

$$Valor\ \Pr edictivoPositivo(VPP) = \frac{a}{a+b}$$

$$Valor\ \Pr edictivoNegativo(VPN) = \frac{d}{c+d}$$

$$\Pr evalencia = \frac{a+c}{a+b+c+d}$$

La *especificidad* de una prueba es la probabilidad de que un sujeto sano sea clasificado correctamente. Una prueba será tanto más específica cuanto menos falsos positivos (FP) proporcione. En general, las pruebas de confirmación diagnóstica deben tener una elevada especificidad, así como en la identificación de enfermedades graves no tratables, en las que un resultado falso positivo puede producir serios trastornos al individuo.

El *valor predictivo positivo* de una prueba diagnóstica es la probabilidad de que un sujeto con un resultado positivo sea un individuo enfermo. El *valor predictivo negativo* es la probabilidad de que un sujeto con un resultado negativo en la prueba sea un individuo sano.

La *seguridad* de una prueba diagnóstica se valora mediante el valor predictivo positivo y negativo, ya que nos dice en qué grado los resultados positivos y negativos de la prueba son correctos. Los clínicos necesitan saber, antes de actuar y tratar, con qué seguridad el paciente tiene la enfermedad cuando la prueba diagnóstica es positiva.

2.1.1. La influencia de la prevalencia en los resultados de una prueba diagnóstica.

Una prueba diagnóstica de calidad debe tener una sensibilidad y especificidad altas (próximas al 100%), pero para garantizar buenos resultados también deben tenerse en cuenta las condiciones de aplicación de la prueba. La prevalencia de enfermedad en la población en la que se aplica el test, aunque no modifica la sensibilidad y especificidad puede modificar los valores predictivos positivos y negativos convirtiendo a una prueba de calidad en una intervención diagnóstica inútil. A continuación se expone un ejemplo.

Si se aplica una prueba con una alta sensibilidad y especificidad, como el test de ELISA para la detección del virus del SIDA, en una población con una prevalencia de seropositividad al VIH muy baja (por ejemplo la población que acude a un determinado centro comercial), se examinará a muchos sujetos sin la enfermedad dando lugar a varios falsos positivos. En la tabla 11-2 se muestran los resultados de la aplicación del test ELISA, en una población con una prevalencia de 1 por mil.

Tabla 11-2. Resultados del test ELISA en una población con baja prevalencia de VIH +.

		VIH +		
		+	-	
ELISA	+	19 VP	200 FP	219
	-	1 FN	19.780 VN	19.781
		20	19.980	20.000

$$Sensibilidad = \frac{19}{19+1} = 0{,}95 \qquad Especificidad = \frac{19780}{19780+200} = 0{,}99$$

$$\Pr evalencia = \frac{20}{19+200+1+19780} = \frac{20}{20.000} = 0{,}001$$

$$VPP = \frac{19}{19+200} = 0{,}086 \qquad VPN = \frac{19780}{1+19780} = 0{,}99$$

Si por el contrario, se aplica el mismo test en una población con alta prevalencia de la enfermedad (39%) aumenta el valor predictivo positivo (tabla 11-3).

Tabla 11-3. Resultados del test ELISA en una población con alta prevalencia de VIH +.

		VIH +		
		+	-	
ELISA	+	63 VP	1 FP	64
	-	3 FN	99 VN	102
		66	100	166

$$Sensibilidad = \frac{63}{3+63} = 0{,}95 \qquad Especificidad = \frac{99}{1+99} = 0{,}99$$

$$\Pr evalencia = \frac{66}{63+1+3+99} = \frac{66}{166} = 0{,}39$$

$$VPP = \frac{63}{1+63} = 0{,}98 \qquad VPN = \frac{99}{3+99} = 0{,}97$$

2.1.2. Las razones de máxima verosimilitud.

Un indicador útil para discriminar entre pruebas diagnósticas por su valor informativo es la "likelihood ratio" que puede traducirse como razón de máxima verosimilitud. Este indicador no tiene un umbral que establezca cuando la prueba diagnóstica tiene poco valor, pero puede ser útil para conocer los resultados que aportará la realización de la prueba.

La razón de máxima verosimilitud del resultado positivo de una prueba diagnóstica se calcula mediante la siguiente fórmula:

$$LikelihoodRatio = \frac{sensibilidad}{1 - especificidad}$$

Cuanto más elevado es el resultado de la fórmula anterior mejor es la prueba. Por ejemplo, se ha estimado que una tomografía computerizada abdominal para el diagnóstico de una lesión esplénica tiene una sensibilidad del 95% y una especificidad del 96%, resultando una razón de máxima verosimilitud de un resultado positivo de 23,8. Sin embargo, los termogramas para la detección del cáncer de mama en la mujer tienen una sensibilidad del 61% y una especificidad del 74%, dando una razón de máxima verosimilitud de un resultado positivo baja (2,35).

La representación gráfica del numerador y denominador de la razón de máxima verosimilitud puede servir de ayuda para discriminar cuál es el punto que ofrece mejores resultados. El siguiente apartado se dedica a la curva ROC, gráfico estándar que representa los puntos de equilibrio entre sensibilidad y especificidad.

La razón de máxima verosimilitud del resultado negativo de una prueba se calcula mediante la siguiente fórmula

$$LikelihoodRatio = \frac{1 - sensibilidad}{especificidad}$$

En este caso, cuanto menor sea el resultado de la Likelihood Ratio mejor es la prueba diagnóstica.

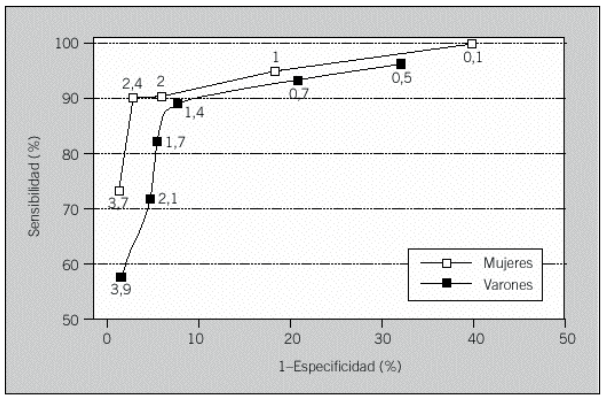
2.1.3. El umbral de positividad de una prueba. La curva de características operativas (curva ROC).

La aplicación de una prueba diagnóstica implica tomar una serie de decisiones que tendrán consecuencias en su evaluación. Un test no aporta la misma información diagnóstica si el umbral para considerar la prueba positiva es bajo que si es muy exigente. La glucemia como prueba diagnóstica de diabetes presenta distintos valores de sensibilidad y especificidad según el umbral diagnóstico sea 100 mg/dl ó 180 mg/dl. Si se considera este último umbral, ciertamente casi todas las personas con una prueba positiva (>180 mg/dl.) estarán enfermas. Sin embargo, muchas otras personas con diabetes no serán calificadas como enfermas con una prueba tan rigurosa. La prueba sería entonces muy específica.

Situándose en el otro extremo, y considerando que todos los pacientes con una glucemia mayor de 70 mg/dl. fueran diagnosticados como diabéticos, muy pocas personas con la enfermedad dejarían de ser diagnosticadas, pero la mayoría de la gente normal sería etiquetada como diabética. La prueba sería entonces muy sensible pero poco específica. Este balance entre sensibilidad y especificidad debe considerarse a la hora de definir el umbral adecuado para considerar positiva la prueba. La construcción de una curva ROC (Receiver Operator Characteristic Curve) es una herramienta muy útil en esa decisión.

Esta curva se construye con los diferentes resultados que se obtienen en una prueba diagnóstica al colocar el umbral de positividad en diferentes puntos. En la curva se representan en un gráfico los pares sensibilidad y 1-especificidad, que en definitiva son las razones de máxima verosimilitud de un resultado positivo. El eje de abscisas crece conforme disminuye la especificidad y el eje de ordenadas crece conforme aumenta la sensibilidad. En la figura 11-1 se muestra un ejemplo de curva ROC para una prueba diagnóstica de nefropatía diabética, el cociente albúmina/creatinina en orina.

Figura 11-1. Curva ROC del cociente albúmina/creatinina.



El eje de abscisas (1-especificidad) corresponde a la tasa de falsos positivos y el eje de ordenadas (sensibilidad) corresponde a la tasa de verdaderos positivos. Los mejores resultados corresponden a los puntos más cercanos al ángulo superior izquierdo, en el que hay una alta tasa de verdaderos positivos y una baja tasa de falsos positivos. La curva con peor resultado sería una línea diagonal que uniera el extremo inferior izquierdo (valor 0-0) con el superior derecho (valor 100-100), indicando que siempre hay los mismos resultados positivos verdaderos que falsos. En ese caso, la razón de máxima verosimilitud de resultado positivo es 1 en todos los puntos de la curva.

La figura 11-1 muestra los diferentes pares de datos sensibilidad/1-especificidad del Cociente Albuminuria/Creatinina para cada uno de los puntos de corte establecidos. El punto de corte o umbral de positividad que más se aproxima al extremo superior izquierdo es el que más área deja debajo de la curva, y corresponde con el de mayor razón de máxima verosimilitud. En el caso de las mujeres el mejor punto de corte es 2,4 g/mol, y corresponde con una sensibilidad de 90% y una especificidad de 97%. En el caso de los hombres es 1,4 g/mol (sensibilidad 89% y especificidad 92%).

Además de poder realizar un análisis cualitativo visualmente, las curvas ROC permiten un análisis cuantitativo preciso para la evaluación de la exactitud de una prueba diagnóstica. El punto que deja mayor área debajo de la curva es el de mayor razón de máxima verosimilitud, y por lo tanto en el que se ofrece una mayor exactitud global.

Visualmente se puede discriminar cuál es el punto que se aproxima más al extremo superior izquierdo, pero para mayor exactitud se calcula el área debajo de la curva. Ésta se interpreta como la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población objeto de estudio, al aplicarles la prueba diagnóstica en cuestión. Es decir, la probabilidad de que el resultado de la prueba resulte más anormal en el sujeto enfermo que en el sano.

El área debajo de la curva es una medida global de la exactitud que permite la comparación estadística entre pruebas diagnósticas. Los valores están comprendidos entre 0,5, el peor resultado posible y 1, el mejor. Un área debajo de la curva de 0,72 significa que un individuo seleccionado aleatoriamente del grupo de enfermos obtendrá el 72% de las veces un valor mayor en la prueba diagnóstica que un individuo elegido al azar del grupo no enfermo.

2.2. Fiabilidad de una prueba diagnóstica.

Los resultados de las mediciones y observaciones clínicas están influidos por multitud de factores como el medio ambiente en el que se realiza la medición o la persona que realiza la observación. Hay un cierto grado de variabilidad o error intrínseco en cualquier procedimiento de medición, particularmente cuando el componente principal del proceso es la apreciación subjetiva de un observador. La variabilidad raramente puede ser eliminada totalmente, pero el conocimiento de sus causas y su evaluación cuantitativa es necesario para garantizar que la información que proporciona la prueba es de calidad.

El grado de variabilidad puede ser evaluado determinando su validez o exactitud, aspecto abordado anteriormente, y además mediante la evaluación de la fiabilidad. La *fiabilidad* es la capacidad de una prueba diagnóstica de obtener resultados comparables si se utiliza en condiciones idénticas. Las pruebas diagnósticas deben tener una elevada fiabilidad, de tal forma que haya consistencia o ausencia de desacuerdo en repetidas observaciones o mediciones. La evaluación de la fiabilidad de una prueba diagnóstica se realiza valorando cada una de las fuentes de variabilidad.

Las diferentes fuentes de variación en las observaciones clínicas pueden resumirse como sigue:

— La *variabilidad intraobservador:* el mismo profesional sanitario puede obtener diferentes resultados al repetir una medición en el mismo paciente debido al cambio en el estado físico del profesional, en el grado de atención u otras condiciones ambientales o psicológicas (por ejemplo: presión por parte de los familiares para emitir rápidamente el diagnóstico).

— La *variabilidad interobservador:* las mediciones realizadas por diferentes profesionales suelen ser diferentes, lo cual puede estar motivado por la formación del profesional, la experiencia, la motivación, la agudeza visual o auditiva y otras características personales que pueden afectar en los resultados de las observaciones.

— La *variabilidad biológica* inherente a la característica que se desea medir. Por ejemplo, la tensión arterial está sujeta a cambios en dependencia de la situación en que se encuentre el paciente en el momento de la toma (ingesta previa, estrés, condición física, etc.).

— La *variabilidad del instrumento de medida.* El propio aparato de medición está sometido a un grado de variación diverso, que puede estar motivado por la calidad de la calibración, las unidades en que se mide o la influencia sobre el propio instrumento de factores ambientales como el nivel de ruido o la iluminación.

— La *variabilidad aleatoria.* Es el grado de variación que no puede atribuirse a los factores anteriores. La repetición de una prueba suele aumentar la fiabilidad al disminuir el error aleatorio. Otro fenómeno reseñable es el denominado "regresión a la media" que es el efecto de disminución de los valores extremos que se produce al realizar repetidamente una prueba, de tal forma que los valores se van aproximando cada vez más a su valor medio real.

La reducción de la variabilidad hasta valores aceptables debe hacerse actuando sobre las diferentes fuentes de variación identificadas. La estandarización del procedimiento de medición puede contribuir a reducir la variabilidad homogeneizando las condiciones de aplicación del procedimiento diagnóstico o entrenando de forma similar a los distintos observadores. Otra forma de reducir la variabiliadad es repitiendo la medición en varias ocasiones para tomar la media de las mismas como valor verdadero del sujeto estudiado.

La evaluación de la fiabilidad de una prueba diagnóstica se realiza valorando cada una de las fuentes de variabilidad. Si lo que se pretende es evaluar la variabilidad interobservador se compararán la concordancia entre los resultados obtenidos por los distintos profesionales, manteniendo constantes el resto de los factores que pueden influir en los resultados.

En dependencia del carácter de la variable que se mide se utilizará un método de evaluación de la variabilidad o concordancia diferente. Cuando se trate de variables nominales o binarias (enfermo/sano, grupo sanguíneo A/AB/B/O) se utilizará el coeficiente Kappa. Si las variables son ordinales se utilizará el coeficiente Kappa ponderado (leve/moderado/grave). Finalmente, cuando las variables sean cuantitativas se utilizará el coeficiente de correlación intraclases.

2.2.1. El coeficiente de concordancia Kappa.

Para desarrollar el concepto se utilizará un ejemplo ilustrativo basado en la evaluación de una prueba diagnóstica de detección de sangre oculta en heces. Los resultados de la prueba pueden ser negativos (0) o positivos (+). La prueba fue realizada por dos médicos (A y B) en 110 pacientes (tabla 11-4).

Tabla 11-4. Resultados de las observaciones de dos médicos sobre un mismo test de detección de sangre oculta en heces.

	Sangre en heces		
	0	+	TOTAL
Médico A: Nº de lecturas en los pacientes	48	62	110
Médico B: Nº de lecturas de los pacientes	47	63	110

La prueba parece perfectamente fiable al presentar resultados aparentemente concordantes entre los distintos observadores. Sin embargo, no sabemos si las observaciones son las mismas

en los distintos pacientes. Si se describen las lecturas mediante una tabla cruzada (tabla 11-5), se observa la concordancia no es completa. La proporción de acuerdo se calcula mediante el Índice de Acuerdo Global (IAG), sumando los valores de las celdas en la que los observadores han coincidido, dividido por el número total de observaciones.

Tabla 11-5. Lecturas completas de la prueba de sangre oculta en heces.

		Médico A		
		0	+	Total
Médico B	0	37	10	47
	+	11	52	63
	Total	48	62	110

En la tabla 11-5 el $IAG = \frac{37+52}{110} = 0{,}81$.

Lo que significa que hay coincidencia entre los dos médicos en el 81% de los casos.

Sin embargo no sabemos cuál habría sido la coincidencia por azar. El cálculo del coeficiente de concordancia Kappa sirve para conocer el nivel de acuerdo que se obtendría si descontamos el debido solamente al azar. En la tabla 11-6 se procede a calcular cuáles serían los resultados si la única relación entre las conclusiones de los médicos A y B fuese el azar, o lo que es lo mismo si los sucesos fuesen independientes. El resultado es el Índice de Acuerdo atribuible al Azar (IAA) que resulta un 0,51.

Tabla 11-6. Frecuencias esperadas por azar de la prueba de sangre oculta en heces.

		Médico A		
		0	+	Total
Médico B	0	20,51	26,47	47
	+	27,49	35,51	63
	Total	48	62	110

* Se obtiene de la multiplicación de los marginales de la fila y la columna en cuestión, dividido por el número total de observaciones (110).

$$IAA = \frac{20{,}51+35{,}51}{110} = 0{,}51$$

Si al Índice de Acuerdo Global (IAG) le descontamos el Índice de Acuerdo atribuible al Azar (IAA) quedaría sólo el nivel de concordancia que no puede ser atribuida al azar. En nuestro ejemplo sería 0,81-0,51=0,30. Este 30% es el acuerdo observado no debido al azar y constituye el numerador del coeficiente Kappa. Para estimar el denominador se calcula la cantidad máxima de acuerdo que podría alcanzarse no debido al azar, es decir el opuesto del IAA. De este modo, el coeficiente Kappa se expresaría:

$$\text{Kappa} = \frac{IAS - IAA}{1 - IAA}$$

En nuestro ejemplo Kappa = 0,30 / 0,49 = 0,61. Por tanto, el 61% de las observaciones son coincidentes una vez eliminado el componente del azar.

Los resultados del coeficiente Kappa varían desde –1, que corresponde con el mayor desacuerdo, a +1, que corresponde con el acuerdo perfecto. El valor 0 correspondería al acuerdo debido al azar. La fiabilidad se considera escasa si el coeficiente Kappa es menor de 0,4; buena si se sitúa entre 0,4 y 0,75 y excelente si es superior a 0,75.

El coeficiente *Kappa Ponderado* se utiliza cuando las variables son ordinales, de tal forma que haya más de dos categorías de valoración con un orden jerárquico entre ellas. Se calcula de forma análoga al coeficiente Kappa Simple, pero multiplicando el valor de cada casilla por un peso que pondere la magnitud del desacuerdo, ya que no es lo mismo discrepar entre leve y moderado, que entre leve y grave.

2.2.2. El Coeficiente de Correlación Intraclases.

El Coeficiente de Correlación Intraclases (CCI) se utiliza para valorar la coincidencia cuando las variables medidas son cuantitativas. Es incorrecto utilizar el Coeficiente de Correlación de Pearson (CCP) u otros índices de tendencia para evaluar la fiabilidad.

El siguiente ejemplo ilustra las diferencias entre el CCP y el CCI. En el ejemplo dos médicos toman la tensión arterial sistólica de forma consecutiva y con el mismo esfigmomanómetro a 10 pacientes. Como puede apreciarse en la tabla 11-7, las mediciones de los dos médicos son diferentes. Los valores registrados por el médico B son en todos los casos superiores a los registrados por el médico A, y además la diferencia es de igual magnitud casi siempre (en 8 de las 10 ocasiones la diferencia es de 5 mm de Hg, y en las dos restantes la diferencia es de 10 mm de Hg). Las diferencias son sistemáticas, van siempre en la misma dirección y son de similar magnitud. En este caso el CCP detecta una alta correlación, concretamente de 0,95. Sin embargo, el CCI sí que detecta que las diferencias son sistemáticas y el resultado refleja una concordancia de 0,64.

Tabla 11-7. Medición de la tensión arterial sistólica.

Paciente	Médico A (mm Hg)	Médico B (mm Hg)	Diferencia (B-A)
1	135	140	5
2	140	145	5
3	130	135	5
4	145	150	5
5	140	145	5
6	150	160	10
7	140	145	5
8	135	140	5
9	140	145	5
10	135	145	10
Media	139	145	6
Desviación Estándar	5,68	6,67	2,11

El cálculo del CCI se basa en un modelo de análisis de la varianza de medidas repetidas. Se descompone la variación total observada (varianza) en los siguientes componentes: la variabilidad debida a las diferencias entre pacientes o individuos observados (Varianza entre-individuos = σ_E); la varianza intra-individuos (σ_I), en este caso debida a diferencias en mediciones repetidas realizadas por distintos observadores, y por último la varianza residual (σ_R), que es la parte de variabilidad inexplicable y que atribuimos al azar. El CCI se define como el cociente entre la variabilidad entre-individuos y la variabilidad total

$$CCI = \frac{\sigma_E}{\sigma_E + \sigma_I + \sigma_R}$$

Cuando una prueba diagnóstica sea muy fiable o cuando la concordancia entre dos observadores sea muy alta, los resultados de las mediciones repetidas sobre un mismo individuo serán muy parecidas y, en consecuencia, la varianza intra-individuos (σ_I) será muy baja y el valor del CCI será alto. Cuanto más se aproxime a 0 el valor de la varianza intra-individuos, más se acercará a 1 el valor del CCI y viceversa. Puesto que es una proporción, los valores del CCI pueden variar de 0 a 1, donde 0 indica ausencia de concordancia, y 1 concordancia o fiabilidad absoluta. Como norma general, valores inferiores a 0,4 indican escasa fiabilidad, entre 0,4 y 0,75 la fiabilidad es buena, y valores superiores a 0,75 reflejan un grado excelente de fiabilidad.

3. ANÁLISIS DE DECISIONES CLÍNICAS.

En cualquier contacto clínico el médico o profesional sanitario se enfrenta a la responsabilidad de decidir. Un razonamiento habitual en la práctica es el *heurístico*. Éste se utiliza cuando se piensa y se valora una situación integrando los conocimientos y las experiencias adquiridas de forma espontánea. Sin embargo, existen técnicas, agrupadas bajo el epígrafe de Técnicas de Análisis de Decisión, que estructuran y facilitan la toma de decisiones mediante mecanismos explícitos que permiten evaluar y hacer más racional la actividad clínica.

El *Análisis de Decisión* puede definirse como una herramienta cuantitativa que evalúa el valor relativo de cada una de las opciones existentes en una decisión. Divide los problemas complejos en partes más sencillas, los analiza en detalle y los combina de nuevo de manera que se pueda identificar la mejor estrategia. Trata de evitar los errores producidos al establecer juicios clínicos basados fundamentalmente en la intuición clínica o en la experiencia personal.

El análisis de decisión es un procedimiento de apoyo a la toma de decisiones, basado en la utilidad esperada de cada opción desde una determinada perspectiva. A la vez, constituye un método sistemático de identificación de opciones, cuantificación de expectativas o probabilidades, asignación de valores a resultados y comparación entre opciones. En medicina se utiliza tanto para facilitar la toma de decisiones en la práctica clínica, ante pacientes concretos, como para diseñar estrategias globales aplicables a un amplio grupo de pacientes o a la política sanitaria.

El abordaje de un problema mediante una técnica de análisis de decisión está indicado cuando se dan las siguientes circunstancias:

— Existe un conjunto de alternativas reales bien definidas entre las que escoger.
— Se detecta incertidumbre sobre las consecuencias de las alternativas.
— La persona o personas que deciden tienen preferencias acerca de los posibles resultados.

— Existen una o más restricciones que limitan el conjunto de alternativas. Por ejemplo, el coste de oportunidad de vacunar a toda la población escolar de hepatitis B es tan grande que nos impide realizar otras actuaciones sanitarias con una mejor relación coste-efectividad.

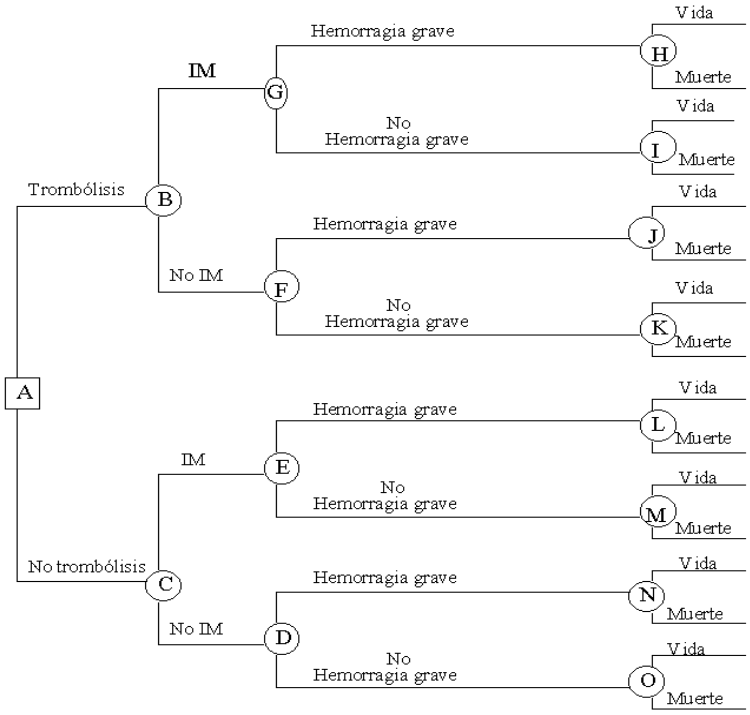
— El problema precisa que se tome una decisión.

En algunos estudios que abordan una situación clínica complicada, el análisis de decisión revela la necesidad de obtener nuevas evidencias científicas que permitan tomar decisiones con base científica. En otros casos puede indicar que las diferencias entre los resultados de las distintas estrategias son muy pequeñas, lo que suele deberse a errores del método. Incluso cuando resulta potencialmente útil, puede no ser factible completar los cálculos o el análisis de sensibilidad precisos por limitaciones de tiempo o de recursos. No obstante, el valor del enfoque analítico del proceso de decisión radica en que integra los datos disponibles, obliga a un razonamiento riguroso y expone las áreas de incertidumbre o ignorancia.

3.1. Los Árboles de Decisión.

Los Árboles de Decisión son las técnicas más usadas entre la amplia gama de técnicas de análisis de decisión existentes. Son útiles para explicitar la secuencia temporal y lógica de las diferentes alternativas de un problema, de tal manera que sea fácil comprender todas sus posibles decisiones y consecuencias. La figura 11-2 muestra un ejemplo de árbol de decisión para decidir si tratar con un determinado fármaco trombolítico intravenoso a los pacientes en los que se sospecha infarto agudo de miocardio.

Figura 11-2. Árbol de decisión sobre la trombolisis en pacientes con sospecha de infarto agudo de miocardio.



Fuente de la figura 11-2: Mundet X, et al. Utilidad del cociente albúmina/creatinina en el diagnóstico de la nefropatía en pacientes con diabetes mellitus tipo 2. Med Clin (Barc) 2001; 116: 732-733. Imagen reproducida con autorización del titular de los derechos de explotación de la misma.

La estructura y símbolos (cuadrado y círculo) de los árboles de decisión están estandarizados. Utilizando como marco el ejemplo de la figura 11-2 se explican a continuación

los significados y la forma de elaborar e interpretar los árboles de decisión. El cuadrado o "núcleo" marcado con la letra A indica la decisión que el médico debe tomar. Los círculos, marcados de B a O, indican los diferentes resultados. Cada uno de ellos proviene de una rama en la que se debe escribir la probabilidad conocida que tiene de producirse ese resultado. Cada una de las evoluciones finales posibles recibirá una utilidad o preferencia, o sea, una valoración cuantitativa de lo apreciado que resulta ese resultado en relación con los demás. El valor de cada resultado final se obtiene multiplicando todas las probabilidades de cada una de las ramas por la utilidad asignada a ese final de rama. Por último, para obtener el valor final de decidir tratar con trombolisis frente al valor de no hacerlo, se suman los valores relativos de todas las puntas de rama resultado de esas dos decisiones, y posteriormente se elegirá la opción que muestre el mayor valor. La tabla 11-8 describe los pasos del análisis basado en árboles de decisión.

Tabla 11-8. Fases del diseño de un análisis basado en un árbol de decisión.

1.	IDENTIFICAR Y LIMITAR EL PROBLEMA, de tal manera que se establezcan todas las alternativas de decisión y los resultados a que pueden dar lugar. Se deben incluir los aspectos más relevantes que puedan determinar los resultados y la utilidad de los cursos de acción, definir claramente la perspectiva del análisis, identificar todas las alternativas y el horizonte temporal de análisis.
2.	CONSTRUIR UN ÁRBOL DE DECISIÓN QUE ESTRUCTURE EL PROBLEMA. Selección de las opciones, nodos de decisión y nodos finales. Debe mantenerse el equilibrio entre complejidad y simplicidad, logrando que los elementos esenciales, tanto en lo que se refiere a opciones de acción como a valores estén incluidos en el modelo.
3.	ESTIMACIÓN DE PROBABILIDADES Y ASIGNACIÓN A LOS CURSOS DE ACCIÓN. Recopilar información mediante una revisión sistemática de la literatura, consulta a expertos o búsqueda de datos primarios, para estimar las probabilidades que van a ser asignadas a cada uno de los componentes del árbol. La probabilidad final de que ocurra el suceso definido por una rama se halla multiplicando todas las probabilidades de todos los trayectos de rama necesarios para llegar a ese resultado clínico final.
4.	ASIGNACIÓN DE VALORES Y UTILIDADES A LAS CONSECUENCIAS O RESULTADOS FINALES. Los resultados clínicos finales pueden expresarse de diversas formas: años de vida, años de vida ajustados por calidad, esperanza de vida ajustada por calidad, casos de enfermedad evitados o complicaciones evitadas, o utilidades. La *utilidad* es una medida de la preferencia relativa, perspectiva, que tiene alguien respecto de un determinado resultado. Se expresa en una escala entre 0 y 1, de tal manera que al resultado de menor utilidad, la muerte sería el caso extremo, se le asigna el valor 0; y al de mayor utilidad, la salud plena sería el otro extremo, el valor 1.
5.	INTERPRETACIÓN DE LOS RESULTADOS Y VALORACIÓN DE LA INCERTIDUMBRE. Conociendo la probabilidad de que ocurra cada suceso, o rama del árbol, y su utilidad se dispone de todos los datos para interpretar los resultados y poder elegir el curso de acción más favorable. Por último, es aconsejable realizar un análisis de sensibilidad que nos confirme la estabilidad de nuestros resultados en diferentes circunstancias.

Respecto al diseño de árboles de decisión se han realizado recomendaciones generales para evitar errores e incorrecciones en su construcción, entre las que se incluyen las siguientes:

— Procurar el balance entre riesgos y beneficios de la intervención en las distintas ramas.

— Situar preferentemente dos ramas en cada nodo de decisión.
— Las ramas deben incorporar información sobre las probabilidades de ocurrencia de cada una de las opciones en que se dividen. Esta información debe estar basada en la evidencia científica.
— Asegurar la simetría.
— Incluir todas las condiciones determinantes que pueden afectar a los resultados.

Las utilidades pueden ser estimadas por diversos procedimientos: juicio de expertos; opinión de las personas que realizan el estudio; búsqueda en la literatura científica o estudio ad hoc.

El árbol de decisión es la técnica principal y más empleada en los análisis de decisión. Otras técnicas son el análisis de umbrales de decisión y el proceso de Markov (que se describen a continuación); los modelos de simulación de Monte-Carlo, y los métodos de medición de preferencias entre estados de salud standard gamble, time trade-off y category scaling.

El *análisis de umbrales de decisión* es un tipo de análisis de decisión que utiliza la teoría de Bayes para valorar, por ejemplo, la conveniencia de realizar una prueba diagnóstica a un paciente. Parte del concepto de que realizar un test a un paciente es inútil si no sirve para modificar la estrategia terapéutica o el manejo clínico global de ese paciente. El umbral es el nivel de probabilidad en el que el beneficio previsible de tomar una decisión iguala al de no tomarla o al de tomar una alternativa.

Las *cadenas de Markov, o el proceso de Markov,* son un método cuantitativo dinámico que toma en consideración la probabilidad de que un paciente cambie de un estado de salud a otro en un determinado período de tiempo, es decir, la probabilidad de transición para ese determinado estado. En este modelo es necesario definir por adelantado cuáles son los estados de salud de interés y especificar todas las posibles transiciones o caminos entre estados.

4. EL CASO DE LA EVALUACIÓN DE UNA PRUEBA DIAGNÓSTICA APLICADA EN CRIBADO POBLACIONAL. EL CRIBADO POBLACIONAL DE CÁNCER DE MAMA MEDIANTE MAMOGRAFÍA.

4.1. Conceptos básicos sobre cribado poblacional

El cribado es un proceso de identificación y selección de sujetos con una enfermedad o característica definida en una población. Si se trata de una enfermedad, en ocasiones, el cribado proporciona un diagnóstico precoz de la misma.

El cribado puede ser organizado y sistemático, dirigido a toda la población, habitualmente mediante un procedimiento de citación y recitación basado en listados exhaustivos de la misma (Ej.: Cribado poblacional de cáncer de mama mediante mamografía en mujeres de 50-65 años). Puede dirigirse a subgrupos específicos, como cribado prescriptivo en grupos de alto riesgo (Ej.: Endoscopia en personas con historia familiar de poliposis intestinal). O puede realizarse de forma oportunista, también llamado búsqueda de casos (Ej.: Cribado de fumadores en atención primaria independientemente del motivo de consulta). El cribado puede ser ocasional o periódico, dirigirse a la identificación de una o varias condiciones (enfermedades, factores de riesgo, características), utilizar una sola prueba o un conjunto de pruebas secuenciales.

El objetivo del cribado poblacional de una enfermedad es reducir la carga de morbilidad y mortalidad asociada a la misma enfermedad. Su potencial utilidad se basa en el hecho demostrable de que en la historia natural de dicha enfermedad existe una oportunidad (mediante la aplicación

de una prueba diagnóstica valida, reproducible sencilla, aceptable) de adelantarnos en el momento del diagnóstico, y este hecho permite que la aplicación del tratamiento disponible ofrezca un curso más favorable de la historia natural de la enfermedad (menor morbilidad, mayor supervivencia, mejora de la función o calidad de vida).

El conjunto de factores que tienen que concurrir para que desde una perspectiva social esté justificada la utilización de un prueba diagnóstica en cribado (contexto de baja prevalencia) son múltiples. Intervienen factores del conocimiento de la propia historia natural de la enfermedad, de la disponibilidad de una prueba adecuada para su uso en este contexto, de la factibilidad del programa (organización, registro, participación, calidad de todos los aspectos del programa, recursos, costes).

Dada la complejidad, recursos necesarios y efectos adversos potenciales de los programas de cribado poblacional es exigible la disposición de evidencias del más alto nivel de calidad sobre la eficacia, efectividad y eficiencia de la intervención propuesta.

Cuando no se dispone de información suficiente sobre la eficacia, o esta procede de estudios de baja calidad por su diseño (estudios observacionales) las medidas de efecto de las intervenciones de cribado suelen estar sobrevaloradas debido a una serie de sesgos (de participación voluntaria –personas con mejor pronóstico-; de adelanto en el momento del diagnóstico- el cribado permite conocer antes el diagnóstico pero no es capaz de alterar la historia natural de la enfermedad; de duración de la enfermedad- más probable sobre-representación de casos con fase preclínica detectable de mayor duración y mejor pronóstico).

Por estas razones, la decisión de poner en marcha un programa de cribado de una enfermedad o factor de riesgo debe basarse en la mejor evidencia científica disponible (ensayos aleatorios controlados, estudios de efectividad poblacional, estudios de coste-efectividad) y un proceso de evaluación exhaustivo de todos los aspectos que determinan el logro de los objetivos planteados (eficacia, efectividad, eficiencia, aceptabilidad, recursos disponibles, organización, etc.).

4.2. Cribado de cáncer de mama mediante mamografía

El caso del cribado de cáncer de mama mediante mamografía ha sido controvertido en distintos aspectos, sobre todo desde la perspectiva de la salud pública, como intervención no suficientemente focalizada en las posibles beneficiarias de la misma. Las estimaciones del número necesario de mujeres a cribar para reducir una muerte/año obtenidas tanto de los resultados de los ensayos como de las experiencias de algunos programas están en el rango de 5.000 a 10.000.

La extensión de la realización de cribado de cáncer de mama, ha provocado un aumento en el número de mamografías, de pruebas diagnósticas complementarias y de biopsias mamarias de lesiones detectadas. El salto de una prueba diagnóstica de su uso en contexto clínico de alta prevalencia a su utilización poblacional en mujeres asintomáticas (contexto de baja prevalencia) plantea los problemas de manejo de los falsos positivos, el sobre-diagnóstico y el sobre-tratamiento.

Actualmente todas las CCAA españolas cuentan con programas de cribado de cáncer de mama. La cobertura poblacional es variable, dado que muchas CCAA han optado por el previo desarrollo de proyectos piloto o la implantación inicial en poblaciones limitadas de su ámbito territorial. Los programas se dirigen a mujeres de 50 a 65 años, si bien algunos programas incluyen a mujeres de 45 a 65 años. Todos han optado por la mamografía como prueba diagnóstica, en proyección única o doble, con una periodicidad de 2 años.

4.3. Validez diagnóstica de la mamografía de cribado.

La mamografía de cribado se ha consolidado como la prueba más idónea para su utilización en los programas de detección precoz de cáncer de mama. La amplia experiencia en su uso y las continuas mejoras en su perfil de precisión diagnóstica y seguridad han reafirmado su posicionamiento frente a otras pruebas diagnósticas alternativas o complementarias. Su elevada aceptabilidad, sencillez y bajo coste relativo son características que apoyan su uso en amplios grupos de población.

La sensibilidad de la mamografía de cribado a 1 año obtenida en distintos ensayos y programas demostrativos varía entre el 83% y el 95%. La especificidad obtenida es también aceptable y varía entre el 93% y el 99%. El problema fundamental, común a otros ámbitos de aplicación de pruebas diagnósticas de aceptable validez en contextos de baja prevalencia (en el caso del cáncer de mama próxima a 3 por 1000), es el relativamente bajo valor predictivo positivo de la mamografía inicial de cribado. En distintos ensayos y programas demostrativos el valor predictivo de un resultado positivo, muy asociado a la prevalencia del cáncer de mama y a la especificidad de la mamografía, se sitúa entre un 5% y un 20%. Lo que supone la realización de pruebas diagnóstico-terapéuticas innecesarias en el 80 al 95% de las mujeres a las que se da un resultado positivo en la mamografía.

4.4. Eficacia de la intervención

La evidencia científica proporcionada por los ensayos controlados y aleatorios realizados desde los años sesenta a los ochenta en Europa, Canadá y los Estados Unidos permite esperar que los programas de cribado contribuyan a reducir el riesgo absoluto de mortalidad por cáncer de mama en las mujeres que inician la participación en los programas con 50 a 64 años de edad en un 20% a un 30%. En este grupo de edad se ha mostrado en los ensayos la obtención de mayor potencial de beneficio (Tabla 11-9). Estos ensayos contaron con la participación y seguimiento de más de medio millón de mujeres comprendidas entre 40 y 74 años, en las que se evaluó el efecto de distintas intervenciones (una o dos proyecciones mamográficas, con o sin examen clínico y con periodicidad variable entre 12 meses y 3 años). Se trata de ensayos de campo sobre el efecto de una intervención preventiva cuyo resultado final es la mortalidad por cáncer de mama. Los sujetos incluidos en los estudios son representativos de la población a la que teóricamente se dirigiría la intervención, los criterios de inclusión mínimamente restrictivos y los protocolos de aplicación reproducen lo que serían componentes de los programas de cribado, facilitando la generalizabilidad de los resultados.

Tabla 11-9. Eficacia de la mamografía de cribado para reducir la mortalidad por cáncer de mama. Riesgo relativo de mortalidad por cáncer de mama en grupo de intervención comparado con grupo control.

Estudio	Riesgo Relativo (IC 95%)
HIP de Nueva York	0,79 (0,61-0,95)
Edimburgo	0,87 (0,70-1,08)
Östergotland (Two-county)	0,76 (0,61-0,95)
Kopparberg (Two-county)	0,58 (0,45-0,76)
Estocolmo	0,73 (0,50-1,06)
Malmö	0,96 (0,68-1,35)
Göteborg	0,55 (0,31-0,95)
Canada	1,08 (0,84-1,40)

Numerosos meta-análisis de los ensayos disponibles fueron publicados en los años noventa confirmando la evidencia de efecto de reducción de mortalidad en el conjunto de edades consideradas, siendo más relevante el efecto en las mujeres que inician la participación en los programas a los 50 años. Por el contrario, en las mujeres que inician el cribado a los 40-49 años no se puede descartar la hipótesis de efecto nulo y se concluye que el efecto no puede ser detectado o es muy limitado, y claramente inferior al observado en mujeres postmenopaúsicas.

La forma de presentación y comunicación de los resultados de los ensayos ha contribuido en cierta medida a una percepción sobredimensionada del efecto del cribado. Se han utilizado las razones de la mortalidad acumulada por cáncer de mama observadas en el grupo de intervención y el grupo control, medidas al final del periodo de seguimiento y expresadas en términos relativos (riesgo relativo). Así un riesgo relativo de 0.79 se expresa como una reducción de la mortalidad por cáncer de mama de un 21% atribuible a la intervención de cribado. Los resultados de los ensayos también se pueden presentar como el número necesario de mujeres a cribar para evitar una muerte al año, o la reducción del riesgo absoluto de morir por cáncer de mama en un periodo y grupo de edad determinado. Una mujer de 50 años asintomática podría estar interesada en conocer su riesgo estimado de muerte por cáncer de mama en los próximos 15 años y en cuanto se reduciría potencialmente dicho riesgo absoluto en caso de participar en el programa de cribado. En el caso hipotético de que este riesgo fuera un 5 por 10.000, según las evidencias científicas publicadas se podría esperar que con la intervención este riesgo individual se redujera hasta un valor de 4 por 10.000. La diferencia absoluta sería de un 1 por 10.000.

Recientemente, Gotzsche y Olsen, publicaron en Lancet y en la Cochrane Library los resultados de sendas revisiones sistemáticas sobre la eficacia de la mamografía para reducir la mortalidad por este tumor. En su primer trabajo, los autores daneses analizan la calidad de los ocho ensayos disponibles según el método de randomización y asignación ciega de las participantes al grupo de intervención y control, el uso de asignación individual, la efectividad de la asignación aleatoria en la obtención de grupos comparables, la justificación de exclusiones y la fiabilidad de la medición del resultado final (la asignación de muertes por cáncer de mama en el grupo de estudio y control). El uso excluyente de criterios de calidad lleva a los autores de la revisión a considerar sólo dos de los estudios disponibles (Canadá y Malmö), valorados como de mediana calidad.

La medida sumarizada del efecto de la intervención sobre la mortalidad por cáncer de mama estimada en estos estudios es 1,04 (IC 95%: 0,84-1,27) y sobre la mortalidad total es 0,99 (IC 95%: 0,94 -1,05), en ambos casos interpretadas como nulidad de efecto. Los autores concluyen además que la mamografía de cribado no es eficaz para la reducción de la mortalidad por todas las causas, único resultado que puede ser valorado en los estudios disponibles de forma fiable, y que los programas de cribado existentes no estarían suficientemente justificados. Se puede decir que los autores daneses se sitúan en una posición extrema, no compartida por los propios investigadores autores de los ensayos, que minusvalora la contribución de ensayos relevantes y no considera otra información relevante disponible sobre la eficacia y efectividad de la intervención (estudios observacionales).

Existen un conjunto de factores que condicionarían la variabilidad de los resultados obtenidos por los ensayos y que pueden tener un peso relevante en la sobreestimación o infraestimación del efecto en cada uno de los estudios. Estos factores no estarían exclusivamente relacionados con el diseño y calidad del estudio, aspectos en los que centran la atención los autores daneses y que utilizan como criterios excluyentes en el análisis, sino que estarían asociados a las características diferenciales de las poblaciones estudiadas, el año de inicio del estudio (contexto de incidencia, actualización diagnóstica y terapéutica, conocimiento público

del efecto), las distintas modalidades de intervención, y la variedad en las periodicidades y tipos de análisis desarrollados.

4.5. Efectos adversos

La realización de la mamografía de cribado implica también ciertos riesgos, fundamentalmente punciones, biopsias y otras pruebas diagnósticas. La realización de la mamografía provoca molestias de intensidad variable y un riesgo reducido de cáncer de mama atribuible a la irradiación. Los riesgos más importantes se relacionan con los resultados falsos positivos y su manejo clínico. Los falsos negativos también pueden generar una falsa seguridad que se traduzca en un retraso en el diagnóstico clínico y aparición de cánceres de intervalo. Se ha comprobado que la mamografía de cribado positiva puede afectar psicológicamente a la mujer por la ansiedad que origina, a lo que se unen los posibles efectos adversos de nuevas exploraciones.

La mamografía de cribado y el conjunto de actividades del programa deben marcarse la excelencia y garantizar una alta calidad que permita minimizar los efectos adversos de la aplicación masiva de la prueba. Por otra parte, la mujer debe ser informada de los beneficios e inconvenientes de la realización de la mamografía y debe participar libremente en la decisión de someterse periódicamente a la prueba.

Un porcentaje muy reducido de mamografías de cribado positivas son confirmadas finalmente como cáncer de mama (5-20%). Por lo tanto, el 80-95% de las mujeres con mamografía positiva inician estudios diagnósticos (examen clínico, mamografías adicionales, ecografía, citología, punción, biopsia) que finalmente no confirman el diagnóstico de cáncer de mama. Se ha observado un incremento significativo del riesgo de mastectomías y tumorectomías entre las mujeres a las que se realiza mamografía frente al grupo control (sin mamografía) (1,31 (IC 95%:1,2-1-42)) y también un aumento de la probabilidad de tratamiento con radioterapia (1,24 (1,04-1,49), lo que a su vez se ha asociado a un incremento de la mortalidad cardiovascular.

4.6. Efectividad del cribado poblacional de cáncer de mama

Resulta muy difícil atribuir a la detección precoz obtenida en los programas de cribado los cambios observados en la mortalidad que también están determinados por numerosos factores, entre los que destacan las mejoras en el manejo clínico y la actualización terapéutica. La reciente reducción en las tasas de mortalidad por cáncer de mama en mujeres de 20-69 años observada en el Reino Unido y los EE.UU. en la década de los noventa se atribuye fundamentalmente a la actualización terapéutica (tratamiento sistémico y cirugía).

A pesar de la difusión de los programas de cribado, algunos con cobertura nacional y más de 15 años en desarrollo, es muy limitada la evidencia sobre el impacto de los mismos en la evolución de las tasas de mortalidad por cáncer de mama. Algunas valoraciones de resultados intermedios o finales de programas de amplia cobertura realizados en países como el Reino Unido, Holanda o Finlandia arrojan resultados inicialmente positivos o prometedores. Sin embargo, el efecto observado en otros países es muy limitado (Suecia).

Es necesario reconocer que los programas de cribado están teniendo un efecto global sobre la mejora en la organización de la asistencia al cáncer de mama, concentrando recursos en la atención a este tumor, reduciendo la demora diagnóstica y terapéutica, actualizando los tratamientos, facilitando la estandarización de la práctica clínica, mejorando la calidad de la atención general e incrementando la equidad y accesibilidad al tratamiento más adecuado.

La evidencia científica, aún basada en estudios de alta calidad, rara vez resulta suficientemente clara, consistente y determinante como para dar respuesta a todos los interrogantes e incertidumbres planteados y representar una guía no sujeta, a su vez, a interpretación. La misma evidencia científica ha servido tanto para justificar la aplicación y difusión de programas nacionales como para el rechazo de los mismos a cualquier nivel poblacional en distintos países desarrollados.

El factor más relevante a valorar es si desde la perspectiva social, o de salud pública, los beneficios potenciales de la intervención superan a los costes y riesgos asumidos por el conjunto de las mujeres, en cuánto los superan y en qué condiciones se puede lograr este objetivo (organización, calidad y excelencia de los programas de cribado). Las mujeres convocadas a participar periódicamente en los programas deben conocer de forma razonable la información disponible sobre los beneficios potenciales más relevantes, en términos absolutos (potencial de reducción de su riesgo personal real) y relativos, así como los riesgos y efectos adversos asociados más relevantes y su frecuencia.

5. REFERENCIAS BIBLIOGRÁFICAS

FLETCHER RH, FLETCHER SW, WAGNER EH. *Epidemiología Clínica.* Barcelona: Ediciones Consulta, SA, 1989.

GOTZSCHE PC, OLSEN O. *Is screening for breast cancer with mammography justifiable?* Lancet 2000; 355:129-34.

JENIZEK M. *Epidemiología. La lógica de la medicina moderna.* Barcelona: Masson, SA editores, 1996.

JONNSON H, NYSTROM L, TORMBERG S, LENNER P. *Service screening with mammography of women aged 50-69 years in Sweden: effects on mortality from breast cancer.* J Med Screen 2001; 8:152-60.

MUSHLIN AI, KOUIDES RW, SHAPIRO DE. *Estimating the accuracy of screening mammography: a meta-analysis.* Am J Prev Med 1998;14(2):143-153.

REY CALERO J, HERRUZO CABRERA R, RODRÍGUEZ ARTALEJO F. Fundamentos de Epidemiología Clínica. Madrid: Editorial Síntesis, S.A,; 1996.

RODRÍGUEZ ARTALEJO F, BANEGAS BANEGAS JR, GONZÁLEZ ENRÍQUEZ J, MARTÍN MORENO JM, VILLAR ÁLVAREZ F. *Análisis de decisiones clínicas.* Med Clin (Barc) 1990; 94: 348-354.

SACKETT DL, HAYNES RB, GUYATT G, TUGWELL P. *Epidemiología Clínica. Ciencia básica para la medicina clínica (2ª edición).* Buenos Aires: Editorial Médica Panamericana S.A., 1994.

SACKETT DL, RICHARDSON WS, ROSENBERG W, HAYNES RB. *Medicina Basada en la Evidencia. Cómo ejercer y enseñar la MBE.* Madrid: Churchill-Livingstone, 1997.

XII. REVISIÓN SISTEMÁTICA Y METAANÁLISIS

Miguel Delgado Rodríguez

1. INTRODUCCIÓN

El término metaanálisis precede cronológicamente al concepto de revisión sistemática. Lo acuñó el psicólogo Glass en 1976 para definir un conjunto de técnicas que se utilizan para cuantificar la información contenida en estudios similares que valoran una misma pregunta de investigación. Surge ante la abundancia de información existente, muchas veces con resultados contradictorios, como una técnica eficiente y reproducible, en la que la unidad de análisis no es el individuo sino el estudio de investigación.

Para que la extrapolación de resultados a la población tenga sentido es necesario que se haga sobre el mayor número posible de investigaciones y ello da origen al concepto de revisión sistemática. La revisión sistemática es definida por la colaboración Cochrane como una síntesis de los resultados de varios estudios primarios mediante técnicas que limitan los sesgos y el error aleatorio. En la actualidad el término metaanálisis se reserva para denominar las técnicas estadísticas de combinación de resultados, mientras que en el pasado incluía también la noción de revisión sistemática.

La idea de combinar los resultados de estudios pequeños para llegar a conclusiones más firmes no es reciente. Por ejemplo, Karl Pearson en 1904 agrupó datos de instalaciones militares de Sudáfrica y la India y concluyó que la vacunación contra la fiebre intestinal no era eficaz; Joseph Goldberger, el investigador de la pelagra, en 1907 publicó una síntesis de la frecuencia de infección urinaria en la fiebre tifoidea. Pero no es hasta finales de los años 1970 cuando se empiezan a realizar metaanálisis a gran escala. Inicialmente se consideró que sólo debía aplicarse para la combinación de estudios de intervención, preferiblemente aleatorizados (la colaboración Cochrane mantiene este criterio), aunque su uso se ha extendido a cualquier tipo de diseños.

2. OBJETIVOS

La revisión sistemática se puede aplicar a cualquier pregunta de investigación, ya sea sobre la etiología (por ejemplo, factores del estilo de vida como causa de enfermedad), diagnóstico (por ejemplo, sensibilidad y especificidad de la TAC en lesiones neoplásicas), pronóstico (por ejemplo, valor de la p53 en el cáncer) o de intervención, ya sea de prevención (por ejemplo, eficacia de una vacuna) o de tratamiento (por ejemplo, los fibrinolíticos en el tratamiento del infarto de miocardio). Los objetivos que se pueden perseguir en estos planos son: (a) valorar la consistencia (o su ausencia, la heterogeneidad) entre los diferentes estudios reunidos; (b) obtener un estimador global de la relación entre las variables objeto de la investigación; (c) identificar los subgrupos que son particularmente susceptibles a la exposición que se estudia; y (d) valorar la calidad de cada una de las investigaciones individuales y ofrecer un esquema de la metodología a seguir en futuros estudios.

3. PROTOCOLO

El esquema general de diseño es el siguiente:

1. Hipótesis de trabajo.
2. Revisión sistemática
   a) Selección de la población de estudio: fuentes de información, criterios de búsqueda de la información y criterios de inclusión de una investigación.
   b) Recogida de datos: valoración de la validez de los estudios primarios y extracción de la información relevante de cada estudio.
3. Metaanálisis: métodos estadísticos empleados para la combinación de resultados, valoración de la heterogeneidad y del sesgo de publicación.

4. ORIGEN DE LA HIPÓTESIS

Lo ideal es que se origine con una información diferente de la que luego se utilizará. Por ejemplo, comprobar en ratones que la dieta rica en ajo disminuye el riesgo de cáncer inducido por carcinógenos puede ser el detonante de una revisión sistemática de estudios humanos sobre esa asociación. No obstante, lo habitual es que la hipótesis de una revisión sistemática se origine tras la lectura de varios informes, que luego formarán parte de la población de estudio. Esto constituye una debilidad bajo la lógica formal de la investigación, ya que para verificar una hipótesis no se puede utilizar la misma información con la que se genera. Hacer una revisión sistemática sobre todos los tratamientos de una enfermedad 'para ver lo que sale', en la que no hay una hipótesis a priori, es un estudio exploratorio en el que no se puede sacar nada, como no sea realizar un nuevo ensayo clínico para verificar las ideas generadas por la revisión.

5. SELECCIÓN DE LA POBLACIÓN DE ESTUDIO

5.1. Búsqueda de la información

5.1.1. Estrategias

La población de referencia en una revisión sistemática está configurada por todos los estudios realizados sobre un tema concreto. Existen varias estrategias:

a) Buscar toda la información disponible, publicada y no publicada. Para aproximarse a la no publicada se pueden buscar las tesinas de licenciatura y tesis doctorales, becas concedidas y solicitadas, contratos de investigación, actas de congresos y el contacto con especialistas en el tema. El problema es encontrar bases de datos que contengan esta información con una perspectiva internacional, lo que exige tiempo y dinero. Otro problema es su calidad, que puede ser muy variable, y diferir la de los estudios no publicados de la de los publicados. Frente a estos inconvenientes, esta estrategia presenta la ventaja de que intenta evitar el sesgo de publicación.

b) Utilizar sólo estudios publicados: es lo más frecuente. Ahorra tiempo y dinero con respecto a la opción anterior. Presenta el problema del sesgo de publicación: lo publicado no tiene porqué representar todo lo realizado.

c) Utilizar las bases originales de datos de los estudios que se combinan: cada vez se difunde más su uso. El primer estudio de este tipo se publicó en 1990: 12 grupos se unieron para examinar la hipótesis de dieta y cáncer de mama a través de sus 12 estudios

de casos y controles. No fue exhaustivo, ya que en ese momento existían muchos más estudios. Las ventajas de la combinación de bases de datos son que permite el análisis de exposiciones raras y que se pueden examinar mejor las relaciones de confusión e interacción. Entre sus inconvenientes están que es menos factible que otras estrategias, ya que los autores deben dar los datos (mayor posibilidad de sesgo de selección) y que requiere más tiempo.

5.1.2. Métodos de búsqueda

Los métodos de búsqueda de información que se pueden utilizar son:

a) Uso de repertorios informatizados: El más usado es MEDLINE, que tiene el inconveniente del sesgo hacia las revistas anglosajonas, 55% del total. Sólo cubre un 20% de las revistas publicadas. Se recomienda en el momento actual que se consulten cuantas más bases de datos mejor; no obstante, se siguen viendo revisiones que sólo utilizan Medline. Las estrategias de búsqueda no deben confiar sólo en las palabras clave, también debe buscarse por sinónimos en la opción de búsqueda libre. El principal inconveniente de los repertorios es que no son exhaustivos de todas las revistas que existen y que no incluyen libros (también en ellos aparecen estudios originales) ni literatura gris.

b) Consulta de la bibliografía de cada uno de los artículos localizados: siempre ha de hacerse. Con ello el investigador se beneficia de las revisiones hechas por otros investigadores y facilita la localización de artículos en revistas no recogidas en los repertorios de consulta o errores en la indexación de artículos.

c) Consulta con investigadores que hayan estudiado el problema: ayuda a la identificación de estudios no publicados, con el inconveniente de que la información proporcionada puede ser subjetiva.

d) Revisión de los índices de las revistas: No todas las preguntas de investigación están bien codificadas por palabras clave en los repertorios. Además, y por regla general, los artículos centrados en un tema se suelen publicar en un rango limitado de revistas, como enseña la bibliometría.

f) Otras fuentes: actas de congresos, repertorios de tesis, etc. Esto normalmente se hace para identificar estudios no publicados.

5.1.3. Sesgo de publicación

Por las dificultades enumeradas con anterioridad lo más frecuente es que los investigadores se basen en estudios publicados y aunque intenten sacar a la luz los no publicados, los publicados tienen más probabilidad de ser localizados. El sesgo de publicación se origina cuando lo que se publica no es representativo de lo que se investiga. La trascendencia de este sesgo es grande para investigadores y lectores: si lo que aparece no representa la realidad, se está distorsionando el propio proceso de aprendizaje. Por eso es muy importante saber si este sesgo se presenta y cuáles son las razones que favorecen su presencia.

La culpa de la presencia del sesgo no sólo reside en los evaluadores y comités editoriales de las publicaciones científicas, también reside en los autores, que en muchas ocasiones son los primeros en decidir no enviar sus investigaciones por diversas razones: por la falta de interés de los resultados, por no tener la propiedad de los datos (estudios financiados por la industria, que puede tener interés en que los resultados que no la favorezcan permanezcan en el anonimato), o por problemas de diseño o ejecución de un estudio que impidieron la obtención de resultados válidos.

Los hechos que se conocen que influyen en la presencia del sesgo de publicación son:

a) *Significación estadística del resultado*. Es el más importante y en el que se basan los procedimientos estadísticos para identificar la presencia del sesgo y que se verán al final del tema. Se documentó ya en la década de 1950. Los resultados significativos suelen ser los primeros en publicarse.

b) *Características de la investigación* realizada: algunas de ellas se asocian con una probabilidad diferencial de publicación: los estudios aleatorizados y los multicéntricos se publican más. El tamaño de muestra también influye. Los estudios grandes suponen un mayor esfuerzo en su diseño y ejecución y los autores ponen un mayor interés en que los resultados no permanezcan ignorados, ya sean positivos o negativos.

c) La *fuente de financiación* se asocia con la publicación. El conflicto de interés puede dar lugar a publicaciones engañosas o maquilladas y dar pie a que cierta documentación tenga más probabilidad de ser publicada. Un ejemplo de ello es el de la industria tabaquera con los efectos del tabaquismo pasivo (se recoge la información que resulta menos perjudicial para la industria).

d) El *prestigio de la institución* puede jugar un papel en el sesgo de publicación. Un ejemplo histórico de esto es el artículo de Levin sobre la relación entre el tabaco y el cáncer de pulmón, retenido por el director de JAMA ante la desconfianza surgida por lo novedoso del diseño (uno de los primeros estudios de casos y controles de la historia), hasta que llegó a sus manos un estudio similar (de Wynder y Graham), firmado por un cirujano de prestigio mundial (Graham), conocido por sus innovaciones. Publicó los dos estudios en el mismo número, pero primero el de Wynder y Graham.

e) *Prejuicio*: aquí el proceso editorial tiene más responsabilidad. Por ejemplo, en la relación entre el consumo de cocaína en el embarazo y riesgo fetal se ha comprobado que los trabajos que no hallan un efecto nocivo sobre el feto se rechazan con más frecuencia, a pesar de que su calidad es similar a los restantes: así se está juzgando la metodología por los resultados, si éstos no son creíbles el evaluador piensa que algo debe andar mal en los métodos.

f) *Regresión a la media*: El sesgo de publicación se ha intentado justificar parcialmente como un fenómeno de regresión a la media. Los primeros artículos sobre una misma pregunta de investigación, por su mayor novedad, son más fácilmente aceptados, que los que a continuación se remiten para su publicación, y en los que se puede manifestar con mayor intensidad los aspectos mencionados con anterioridad y que influyen en la publicación.

Para controlar el sesgo de publicación se debe vigilar el proceso de revisión de los manuscritos. Se prefiere la evaluación enmascarada (el revisor desconoce a los autores del artículo), aunque hay dudas acerca de si los autores deben conocer la identidad del revisor o ésta debe mantenerse anónima.

El uso de cuestionarios estructurados en el proceso de evaluación facilita la valoración y la hace más reproducible, pero limita la expresividad de los autores y no todos los aspectos de un estudio pueden contemplarse en un listado de preguntas. No obstante, el metaanálisis ha supuesto la implantación en muchas revistas de las normas CONSORT, para los que quieran publicar un estudio de intervención, y posiblemente ocurrirá lo mismo con las normas STARD para pruebas diagnósticas.

Debieran hacerse esfuerzos para que toda investigación con un mínimo de calidad se publicara, con independencia de la significación de sus resultados, pero el papel se vende caro. Una opción futura es la edición electrónica (internet).

El registro de investigaciones, como la base de ensayos perinatales de Oxford o la Colaboración Cochrane, que intentan monitorizar todo ensayo clínico en marcha, tiene sus limitaciones. Puede haber un conflicto de intereses, por ejemplo, si un organismo privado prefiere que se ignore que un estudio está en marcha (no infrecuente en salud laboral) o cuando se esperan beneficios económicos en el desarrollo de nuevos fármacos. También hay ventajas: valorar el sesgo de publicación, defender los derechos en ser el primero en investigar un tema, conocer a otros que trabajan en temas relacionados, etc.

5.2. Criterios de inclusión

Han de establecerse sin conocer los resultados de los distintos estudios. Así se disminuye el sesgo de selección. Los estudios identificados con la búsqueda constituyen la serie inicial, pero no todos ellos serán incluidos en el análisis. Es necesario definir mediante unos criterios de inclusión los que permanecerán y los que serán desechados.

a) *Idioma*: es muy frecuente restringir el idioma, ya que es imposible dominar todos. Sería aceptable si la finalidad fuera, por ejemplo, la valoración metodológica de un determinado tópico en un cierto ámbito, como un análisis de la calidad de los estudios coste-eficacia en las revistas científicas publicadas en inglés o castellano. Es más discutible la validez externa de metaanálisis sobre asociaciones de causalidad basadas sólo en estudios en inglés, y es frecuente verlo en estudios anglosajones. La calidad superior de las investigaciones en inglés no es un argumento válido. Es más, en el caso de los ensayos clínicos una investigación ha comprobado que las publicaciones en otros idiomas no son peores que las escritas en inglés y que los ensayos clínicos en inglés contienen con más frecuencia resultados estadísticamente significativos. Esto puede suponer un sesgo de selección. Si se quieren limitar los idiomas y mantener una perspectiva poblacional, se puede acotar por lugar de realización del estudio primario, por ejemplo, al mundo occidental. ¿Qué probabilidad hay de que un americano, un español o un francés, por ejemplo, publiquen sus resultados en chino?

b) *Tipo de diseño*: en la combinación de ensayos clínicos es habitual centrarse sólo en los aleatorizados (porque tienen un mayor control de los errores). En metaanálisis de estudios de observación es frecuente eliminar los estudios ecológicos por el efecto impredecible de la falacia ecológica.

c) *Características de la exposición y del efecto*: se han de precisar las que se desean según los objetivos de cada investigación.

d) *Tipo de publicación*: se centra en los informes originales y se eliminan las revisiones, cartas y editoriales. El problema se suscita cuando en una carta se encuentran datos originales; su extensión no permite una correcta evaluación.

e) *Calidad*: este aspecto se discutirá con detalle en el epígrafe siguiente.

6. RECOGIDA DE DATOS

Antes de recoger los datos que se combinarán hay que valorar la calidad de una investigación. Las tres preguntas básicas en la valoración de la calidad son: ¿hasta que punto importa?, ¿cómo se puede medir? y ¿qué utilidad presenta?

6.1. Importancia de la valoración de la calidad

Ayuda a situar la robustez de la información existente. Los ensayos clínicos aleatorizados se prefieren a los no aleatorizados y éstos a los estudios de observación, pero no todos los estudios con un mismo diseño están ejecutados con la misma escrupulosidad. Los estudios de alta calidad reducen la existencia de sesgos y tienen una mayor precisión en el estimador del efecto que calculan. Si se planea hacer metaanálisis, debe hacerse con estudios en los que sea menor la probabilidad de que haya un sesgo.

6.2. Forma de medir la calidad

Hay que confeccionar un cuestionario de evaluación y seguir una estrategia cuidadosa que evite el sesgo del observador, y que podría acabar en un sesgo de inclusión o de puntuación de la calidad, garantizando la reproducibilidad intraobservador y entre observadores. El cuestionario debe aplicarse con enmascaramiento frente a ciertas características: la institución ejecutora, el nombre de los autores (su prestigio puede ejercer un cierto grado de presión) y el sentido de la asociación, ya que su conocimiento podría influir también en la selección. Es recomendable el uso del *scanner* para las secciones de Métodos y Resultados, tachando los epígrafes indicativos de los grupos de pertenencia de los sujetos; uniformiza fácil y rápidamente los formatos reconocibles de las revistas y estandariza la evaluación.

Se recomienda que los evaluadores no sean los que recogen los datos de la magnitud del efecto. La familiarización con los resultados, junto a una posible opinión propia (que puede formarse conforme se lee), pueden contribuir a la aparición de sesgos del observador. Se considera que es suficiente que sean dos evaluadores independientes. Compararán sus conclusiones y si existen discrepancias, deberán intentar solucionarlas entre ellos, y si se mantienen, la decisión final se tomará por un tercer investigador.

Los cuestionarios de evaluación tienen el inconveniente de su validación. No existen cuestionarios ideales para ser aplicados en toda situación. En el caso del ensayo clínico se han publicado muchos cuestionarios y se ha comprobado que no todos los protocolos puntúan igual a los mismos estudios. Otro inconveniente es que con frecuencia se aprecia que no se pueden responder a todas las preguntas de un cuestionario con la información dada en un artículo. En epidemiología de observación no existe ningún cuestionario maestro, y debe adaptarse a cada situación. Con independencia de que se utilice un cuestionario publicado es conveniente que se adapte a la hipótesis sobre la que se trabaja, añadiendo cuestiones sobre los errores particulares que amenazan a la asociación en estudio. También se puede solicitar a los grupos de investigación originales aclaración de aspectos confusos de los Métodos.

6.3. Utilidad de la calidad

Puede tener una triple función:

a) Función cualitativa o discriminar si el resultado final de un estudio está probablemente sesgado. Con ello se valora la inferencia en función de la mejor prueba disponible. Aunque existen algunas dudas, los estudios con sesgos probables no deben ser incluidos. Esta función puede ser esencial en el análisis, por disminuir la heterogeneidad entre estudios. Hay que ser cuidadoso en la exclusión de estudios, puesto que puede introducir un sesgo de selección si los criterios no son objetivos. Por ello el proceso de aceptación tiene que estar enmascarado frente a las variables que luego puedan tener repercusión en la combinación final de las evidencias. Se recomienda que sea realizado por dos investigadores de manera independiente.

b) Incorporación de la calidad como criterio de ponderación: En la actualidad no se recomienda, aunque pueda parecer recomendable dar más peso a los mejores estudios; entre otras razones, porque diferentes cuestionarios dan un orden de calidad diferente a los mismos estudios.

c) Recogida de las características del estudio: tipo de población, criterios de inclusión y exclusión, forma de recogida de datos, técnicas de control de errores, etc. En el caso de que exista heterogeneidad, esta información se puede utilizar para intentar tipificar las características asociadas a un cierto resultado.

7. MÉTODOS DE COMBINACIÓN

Se pueden combinar los resultados publicados o combinar las bases de datos. En este último caso, el análisis estadístico es convencional, salvo que es conveniente incluir en los modelos multivariables una variable que identifique al estudio. Este apartado se dedicará a la combinación de datos procedentes de informes publicados.

Las técnicas estadísticas de metaanálisis generan varios problemas:

— Se puede asumir que el efecto producido por la exposición sea constante de estudio en estudio (modelo de efectos fijos, MEF) o, por el contrario, tenga una distribución al azar a lo largo de los distintos estudios (modelo de efectos aleatorios, MEA).

— Todos los estudios incluidos en un metaanálisis no son formalmente "independientes" ya que los últimos pueden iniciarse tras conocer los resultados de los primeros. Esto sí es un auténtico problema: la investigación actual se apoya en la del pasado.

— Existe una tendencia a encontrar resultados estadísticamente significativos conforme el número de estudios aumenta, aunque no favorezcan la asociación. Por ello, se recomienda que el nivel de significación sea como mínimo del 1%. No obstante, la mayoría de los metaanálisis publicados calculan intervalos de confianza del 95%.

Sea cual sea el procedimiento que se aplique, cada estudio proporcionará el mismo parámetro, θ_i, que luego se combinarán utilizando un criterio de ponderación w_i : $\theta_p = (\Sigma\, w_i\theta_i)/\Sigma\, w_i$.

7.1. Métodos para variables dicotómicas: modelo multiplicativo

Son métodos que se basan en la noción de riesgo relativo (RR). Este modelo es generalmente más adecuado que el aditivo para las relaciones entre una exposición y un efecto. En el caso de que se prefiera aplicar un modelo aditivo, basado en la diferencia de riesgos, se puede aplicar el método del inverso de la varianza que a continuación se describe.

7.1.1. Método del inverso de varianza

Se usa como criterio de ponderación el inverso de la varianza. El procedimiento se resume en la tabla 12-1. Como ejemplo se expone el caso de la *odds ratio* (OR). Cada estudio proporciona una OR y su varianza. Para la combinación de resultados, dado que la OR (al igual que el RR) no sigue una distribución normal, se transformará logarítmicamente:

Tabla 12-1. Resumen de la ponderación por inverso de la varianza (modelo de efectos fijos).

Objetivo	Estimación
Ponderación resumen, θ_p	$\theta_p = \frac{\sum w_i \theta_i}{\sum w_i}$ $w_i = \frac{1}{Var(\theta_i)}$
Error estándar de θ_p	$EE(\theta_p) = \sqrt{\frac{1}{\sum w_i}}$
Intervalo de confianza de θ_p	$\theta_p \times z_{\alpha/2} EE(\theta_p)$
Significación estadística de θ_p	$z = \frac{\theta_p}{EE(\theta_p)}$
Heterogeneidad, Q	$Q = \sum w_i (\theta_i - \theta_p)^2$ χ^2 con k – 1 grados de libertad. k = número de estudios

Si se ponderan RD, qp = RDp, qi = RDi y wi = 1/Var(RDi) (ver tabla 12-2)
Si se ponderan RR, qp = ln(RRp), qi = ln(RRi) y wi = 1/Var(ln RRi) (ver tabla 12-2)

$$\ln ORp = \frac{\sum w_i \times \ln OR_i}{\sum w_i}, \quad w_i = \frac{1}{V(\ln OR_i)}$$

siendo OR_p la odds ratio ponderada. Para calcularlo será necesario realizar la transformación antilogarítmica. La forma de estimar los parámetros crudos se encuentra resumida en la tabla 12-3, basándose en la distribución de la tabla 12-2. En el caso de que se ofrezcan estimaciones ajustadas (controlando por uno o más factores de confusión), los autores suelen proporcionar el intervalo de confianza de la OR. En esta situación, para calcular la varianza de la OR se aplicará la ecuación V (ln OR)= $[(\ln OR_s - \ln OR_i)/3{,}92]^2$, siendo OR_s y OR_i los límites superior e inferior del intervalo de confianza del OR.

Tabla 12-2. Estructura y notación de un estudio *i*

	Efecto sí	Efecto no	Total
Expuestos	a_1	b_i	n_{1i}
No expuestos	c_i	d_i	n_{0i}
Total			n_i

Tabla 12-3. Medidas de magnitud de efecto (con la notación de la tabla 12-1).

Medida	Estimación	Varianza	Intervalo de confianza
Riesgo relativo (RR)	$a_i/n_{1i}/(c_i/n_{0i})$	V (ln RR) = $1/a_i + 1/c_i - 1/n_{1i} - 1/n_{0i}$	$RR \times e^{\pm 1{,}96 \times \sqrt{V(\ln RR)}}$
Odds ratio (OR)	$a_i d_i/(b_i c_i)$	V (ln OR) = $1/a_i + 1/b_i + 1/c_i + 1/d_i$	$OR \times e^{\pm 1{,}96 \times \sqrt{V(\ln OR)}}$
Diferencia de riesgos (DR)	$a_i/n_{1i} - c_i/n_{0i}$	$R_{1i}(1-R_{1i})/n_{1i} + R_{0i}(1-R_{0i})/n_{0i}$	$DR_i \pm 1{,}96\sqrt{V_i}$

La varianza del logaritmo natural de la OR_p es igual al inverso de la suma de los pesos de cada OR, $V(\ln OR_p) = 1/(\sum w_i)$. El intervalo de confianza de la OR_p es $= OR_p \cdot EXP\,[\pm z_{\alpha/2}\sqrt{V}\,(\ln OR_p)]$, siendo $z_{\alpha/2} = 1{,}96$ si se quiere una confianza del 95%.

La significación global del parámetro se puede estimar dividiendo el logaritmo natural del ORp por su error estándar, $\chi_{asoc} = \ln OR_p/\sqrt{[V(\ln OR_p)]}$, que sigue una distribución Normal (0,1).

El grado de homogeneidad se valora mediante el estadístico $Q = \sum w_i(\ln OR_i - \ln OR_p)^2$, que sigue una distribución χ^2 con (k - 1) grados de libertad, siendo k el número de estudios. El punto de corte para detectar heterogeneidad se sitúa en el 10% (p = 0,1) y no en el clásico 5%.

Se puede incorporar explícitamente la variabilidad entre estudios y asumir un MEA, en vez de un MEF, como hasta ahora se ha hecho. El MEA se resume en la tabla 12-4. Hay que calcular la varianza entre estudios:

Tabla 12-4. Resumen de la ponderación por inverso de la varianza (modelo de efectos aleatorios).

Objetivo	Estimación
Ponderación resumen, θ_p	$\theta_p = \frac{\sum w_i^* \theta_i}{\sum w_i^*}$ $w_i^* = \frac{1}{Var(\theta_i) + \tau^2}$ $\tau^2 = \max\left\{0, \frac{Q - K + 1}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}\right\}$ Q, w_i y k como en la tabla 13-3.
Error estándar de θ_p	$EE(\theta_p) = \frac{1}{\sum w_i^*}$
Intervalo de confianza de θ_p	$\theta_p \times z_{\alpha/2} EE(\theta_p)$
Significación estadística de θ_p	$z = \frac{\theta_p}{EE(\theta_p)}$
Heterogeneidad, Q	$Q = \sum w_i^* (\theta_i - \theta_p)^2$ χ^2 con k – 1 grados de libertad

$$\tau^2 = máx.\left\{0, \frac{Q - K + 1}{\sum w_i - \left(\sum w_i^2 / \sum w_i\right)}\right\}$$

El peso de cada investigación vendría dado ahora por: $w^*_i = 1/(V_i + \tau^2)$. Esto supondría que la $OR_p^* = \sum (w^*_i OR_i)/\sum w^*_i$. La varianza y el intervalo de confianza se estiman como en el caso anterior, cuando no se tuvo en cuenta la variabilidad entre estudios. En el presente caso se sustituiría w_i por w^*_i.

El procedimiento anterior se ilustra con un ejemplo sobre el efecto del descenso del colesterol por dieta y fármacos sobre la mortalidad (tabla 12-5). Hay que calcular la varianza de la OR de cada estudio (ver tabla 12-3). Para el estudio LA Veterans es 1/174 + 1/250 + 1/177 + 1/245. En la tabla 12-6 se resumen las varianzas de cada estudio, sus pesos, y los nuevos pesos tras la comprobación de que $\tau^2 = 0{,}0231$. La OR_p, su IC, y la Q valen:

Tabla 12-5. Descenso de la colesterolemia por dieta y medicación y mortalidad: estudios de prevención primaria

Estudio	Tto. experimental / tto. control	Gr. experimental		Grupo control		OR
		Viven (b)	Mueren (a)	Viven (d)	Mueren (c)	
LA Veterans	Dieta especial / convencional	250	174	245	177	0,96 (0,73-1,28)
Minnesota	Dieta especial / convencional	2039	158	2044	153	1,04 (0,82-1,31)
OMS	Clofibrato / placebo	5203	128	5209	87	1,47 (1,11-1,96)
Upjohn	Colestipol / placebo	531	17	519	27	0,62 (0,31-1,19)
Lipids	Colestiramina + dieta / placebo + dieta	1838	68	1829	71	0,95 (0,67-1,36)
Helsinki	Gemfibrozil + dieta / placebo + dieta	2006	45	1988	42	1,06 (0,68-1,66)

Fuente: Muldoon et al. BMJ 1990;301:309-14.

Tabla 12-6. Varianzas y pesos del ejemplo de la tabla 12-6 en la aplicación del método del inverso de la varianza (t2 = 0.0231).

Estudio	V (ln OR)	Peso (1/V)	Peso* [1/(V+τ^2)]
LA Veterans	0.0195	51,3	23,4
Minnesota	0.0138	72,5	27,1
OMS	0.0197	50,8	23,4
Upjohn	0.0997	10,0	8,1
Lipids	0.0299	33,4	18,9
Helsinki	0.0470	21,3	14,3

$$\ln ORp = \frac{51{,}3 \times 0{,}96 + \ldots + 21{,}3 \times 1{,}06}{51{,}3 + \ldots + 21{,}3} = 0{,}0630$$

$$ORp = e^{0{,}063} = 1{,}065$$

$$V(\ln ORp) = \frac{1}{51{,}3 + \ldots + 21{,}3} = 0{,}00419$$

$$IC\ 95\%\ OR = 1{,}045 \times e^{\pm 1{,}96\sqrt{0{,}00419}} = 0{,}94\ a\ 1{,}21$$

$Q = 51{,}3(\ln 0{,}96 - \ln 1{,}045)^2 + ... + 21{,}3(\ln 1{,}06 - \ln 1{,}045)^2 = 9{,}35$, con 5 grados de libertad, $p = 0{,}096$ (hay heterogeneidad entre los estudios).

Como ya se ha mencionado, $\tau^2 > 0$ y se puede aplicar el MEA. Se recalculan los pesos (tabla 12-6). En primer lugar se aprecia que los pesos son menores que en el MEF, lo que se traducirá en una mayor varianza (recuerde, el inverso de la suma de los pesos) y un intervalo de confianza mayor. Se comprueba también que las diferencias en la importancia de los estudios se han amortiguado: con los pesos originales el estudio Minnesota era siete veces más influyente que el de Upjohn (72,5 frente a 10,0), mientras que con la nueva ponderación, es algo superior a tres (27,1 frente a 8,1). Se concede con el MEA una mayor importancia relativa a los estudios pequeños.

Con el MEA, la $OR_p = 1{,}045$ y un intervalo de confianza más amplio que con el MEF, e igualmente no significativo (0,87-1,25). Al descender el valor de los pesos, la prueba de homogeneidad Q se hace más pequeña, $Q = 5{,}36$, $p = 0{,}37$. Es más difícil encontrar heterogeneidad con el MEA.

7.1.2. Método de Mantel-Haenszel

Este procedimiento se recomienda para los estudios experimentales, ya que trabaja con los datos crudos, por lo que el diseño debe utilizar técnicas que eliminen el sesgo de confusión (como la aleatorización). El procedimiento se resume en la tabla 12-7. La OR global (OR_{MH}) se estimará ponderando la OR cada estudio (OR_i) por el coeficiente $c_i b_i / n_i$

Tabla 12-7. Resumen de la ponderación de Mantel- Haenszel de OR.

Objetivo	Estimación
Ponderación resumen, OR_{MH}	$OR_{MH} = \dfrac{\sum_i \frac{a_{1i} b_{0i}}{n_{+i}}}{\sum_i \frac{b_{1i} a_{0i}}{n_{i+}}}$
Error estándar de OR_{MH}	$SE(\ln OR_{MH}) = \sqrt{\dfrac{\sum_i P_i R_i}{2\left(\sum_i R_i\right)^2} + \dfrac{\sum_i (P_i S_i + Q_i R_i)}{2\sum_i R_i \sum_i S_i} + \dfrac{\sum_i Q_i S_i}{2\left(\sum_i S_i\right)^2}}$ $P_i = \dfrac{a_{1i} + b_{0i}}{n_{+i}}; Q_i = \dfrac{a_{0i} + b_{1i}}{n_{+i}} = 1 - P_i; R_i = \dfrac{a_{1i} b_{0i}}{n_{+i}}; S_i = \dfrac{a_{0i} b_{1i}}{n_{+i}}$
Intervalo de confianza de OR_{MH}	$OR_{MH} \times e^{z_{\alpha/2} EE(\ln OR_{MH})}$
Significación estadística de OR_{MH}	$\chi_{MH} = \dfrac{a_1 - \sum m_{1i} n_{1i} / n_{+i}}{\sqrt{\dfrac{m_{1i} m_{0i} n_{1i} n_{0i}}{n_{+i}^2 (n_{+i} - 1)}}}$
Heterogeneidad, Q	$Q = \sum w_i (\ln OR_i - \ln OR_{MH})^2$ χ^2 con k – 1 grados de libertad $w_i = \dfrac{1}{Var(OR_i)} = \dfrac{1}{\frac{1}{a_{1i}} + \frac{1}{a_{0i}} + \frac{1}{b_{i1}} + \frac{1}{b_{0i}}}$

$$OR_{MH} = \frac{\sum_i \frac{a_i d_i}{n_{+i}}}{\sum_i \frac{b_i c_i}{n_{+i}}}$$

La varianza del parámetro resumen será:

$$V(\ln OR_{MH}) = \frac{\sum_i P_i R_i}{2\left(\sum_i R_i\right)^2} + \frac{\sum_i (P_i S_i + Q_i R_i)}{2\sum_i R_i \sum_i S_i} + \frac{\sum_i Q_i S_i}{2\left(\sum_i S_i\right)^2}$$

$$P_i = \frac{a_i + c_i}{n_{+i}}; Q_i = \frac{c_i + b_i}{n_{+i}} = 1 - P_i; R_i = \frac{a_i d_i}{n_{+i}}; S_i = \frac{c_i b_i}{n_{+i}}$$

Intervalo de confianza de OR_{MH} = $OR_{MH} \times EXP\left(\pm 1.96 \times \sqrt{V(\ln OR_{MH}}\right)$

Para la valoración de la heterogeneidad se pueden utilizar varias opciones. El más usual es el estadístico $Q = \sum w_i (\ln OR_i - \ln OR_{MH})^2$, siendo w_i el inverso de la varianza del ln OR. El procedimiento de Mantel-Haenszel se ilustra con los datos del ejemplo de la tabla 12-5.

$$OR_{MH} = \frac{\frac{174 \times 275}{846} + \ldots + \frac{45 \times 1988}{4081}}{\frac{177 \times 250}{846} + \ldots + \frac{42 \times 2006}{4081}} = 1{,}066$$

La fórmula de su varianza es tediosa en su cálculo manual y no se reproduce (está implementada en muchos paquetes estadísticos), da un valor de 0,00414. El IC del 95% es $1.066 \times EXP(\pm 1.96\sqrt{0{,}00414} = 0{,}94\ a\ 1{,}29$, valores muy parecidos a los del procedimiento del inverso de la varianza.

Hay otros procedimientos que se derivan del de Mantel-Haenszel, como el de Peto, muy utilizado. No se comenta porque es un procedimiento sesgado, sobre todo cuando hay desequilibrio en el tamaño de la muestra entre los grupos que se comparan y/o cuando la OR es claramente distinta de la unidad.

7.1.3. Estrategia general de análisis

La elección entre una medida basada en el RR o en la diferencia de riesgos depende de los objetivos de la investigación. En los estudios de observación el método del inverso de la varianza permite combinar parámetros ajustados. El método de Mantel-Haenszel y derivados (método de Peto) obliga a utilizar la distribución cruda de los datos, donde es probable que haya sesgo de confusión. Este hecho supone que estos métodos de combinación se recomienden en general para los estudios experimentales.

Con respecto a la disyuntiva de utilizar el MEF o el MEA, se insiste en que la elección debe especificarse antes y no tras conocer los resultados. Es muy común empezar por un MEF (mucho más restrictivo a la hora de considerar la heterogeneidad, ya que es más fácil detectarla con este procedimiento) y esto basta si no hay heterogeneidad; por el contrario, si hay heterogeneidad se recurre al MEA. A priori se recomienda el MEA cuando el número de estudios es pequeño (< 20).

El MEA asume que se tiene una muestra de los estudios existentes, asunción más realista que tener todo el universo de estudios, como subyace en el MEF. No obstante, conviene saber que la estimación ponderada procedente de un MEA no tiene interpretación poblacional, ya que la distribución del error aleatorio de la muestra no tiene justificación biológica o epidemiológica, al no existir un modelo de distribución conocido, ni un proceso de muestreo subyacentes (un grave inconveniente). El MEA introduce un término de variabilidad adicional y aumenta la varianza del estimador resumen, con un mayor intervalo de confianza. Esto no es una ventaja desdeñable, ya que uno de los inconvenientes que se han señalado en el metaanálisis es la facilidad de obtener resultados estadísticamente significativos. Sin embargo, el mismo coeficiente de variabilidad se añade por igual a cada estudio, lo que concede una mayor importancia relativa a los estudios pequeños frente a los grandes, lo que supone que el MEA se influye más por los errores que afectan a los estudios pequeños, como el sesgo de publicación.

Algunos autores recomiendan que se haga un análisis de sensibilidad, siempre y cuando sea posible, y combinar los resultados de las distintas investigaciones por más de un método. Si los distintos análisis coinciden, esto reforzará la inferencia a realizar de los resultados.

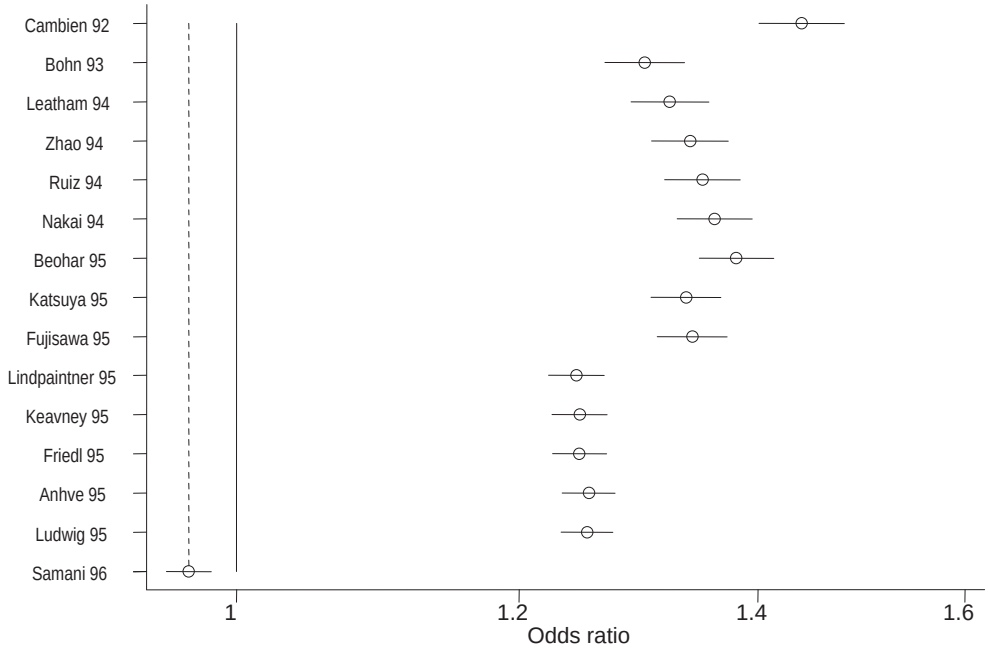
7.2. Métodos para variables continuas

Las variables de efecto continuas se utilizan con frecuencia en epidemiología (por ejemplo, dieta y reducción de las cifras de colesterol sérico). Se puede aplicar el método de inverso de la varianza. En este caso el parámetro que se pondera es una media (por ejemplo, la media de reducción de colesterol tras una intervención dietética) y el peso es el inverso del cuadrado del error estándar de la media (su varianza). También se puede aplicar el análisis de la varianza y diferentes modelos de regresión (ponderados o no); la prueba de homogeneidad es el factor "estudio" o la interacción estudio-exposición.

7.3. Metaanálisis acumulado

Consiste en añadir cada vez un estudio en la estimación del parámetro resumen según una variable cuantitativa que permite la ordenación de las distintas investigaciones. Los resultados se presentan en forma de gráfico, representando el valor del estimador resumen (RR, OR, etc.) y su intervalo de confianza después de cada adición. El más habitual es el realizado según la fecha de publicación: cada vez se añade un estudio más reciente. Sirve para valorar cuál es la contribución que tiene cada estudio sobre el grado de evidencia disponible en ese momento. Un ejemplo se observa en la figura 12-1. También se pueden realizar según otras variables: calidad global de cada estudio, frecuencia del efecto en el grupo de referencia, etc.

Figura 12-1. Ejemplo de metaanálisis acumulado sobre la asociación entre la deleción del gen de la enzima convertidora de la angiotensina e infarto de miocardio.



(Fuente: Samani et al. Circulation 1996; 94:708-12)

8. ESTUDIO DE LA HETEROGENEIDAD

Siempre debe realizarse un análisis de la heterogeneidad. La síntesis de estudios realizados en diferentes lugares y tiempos, sobre distintas poblaciones y con diseños diferentes, con frecuencia da lugar a que los resultados sean estadísticamente diferentes entre sí. Esta heterogeneidad no puede ignorarse. Supóngase que todos los resultados de los estudios individuales son ciertos, sin sesgos, y existe una marcada heterogeneidad; aquí el cálculo de una medida resumen de los diferentes estudios no aporta nada (más bien confunde). El conocimiento de la media entre los diferentes estudios no permite apreciar la riqueza de la variabilidad existente y lo más acertado sería investigar cuáles son las razones para que se produzcan resultados contradictorios. El metaanálisis sólo debe sacar un denominador común cuando de verdad éste existe.

Un inconveniente grave del análisis de la heterogeneidad es que la mayoría de sus pruebas estadísticas no tienen una potencia estadística adecuada; por ello, el no rechazar la hipótesis de homogeneidad no implica que no exista heterogeneidad. En el análisis de la heterogeneidad se debe partir de una reflexión de cuáles pueden ser las variables que pueden ejercer una influencia diferencial: variables de diseño, exposición, efecto, otros factores de riesgo, variables de persona, tiempo y lugar, etc.). A continuación se puede proceder de varias maneras:

a) *Análisis estratificado*: se repite el metaanálisis en cada uno de los estratos. Si desaparece la heterogeneidad es razonable pensar que esa variable influye. El inconveniente que presenta esta estrategia es que puede resultar poco eficiente si hay muchos estratos y el número total de estudios es pequeño.

b) *Metarregresión*: Suele ser más eficiente que la técnica anterior. En ella se utiliza como variable dependiente la magnitud del efecto que se valora con una o varias variables independientes: $\ln OR = \beta_0 + \sum \beta_i x_i$. Las técnicas de metarregresión trabajan con datos agregados y por lo tanto están sometidas a los problemas de la falacia ecológica.

c) *Metaanálisis acumulado*: es menos sensible que las anteriores para la identificación de la heterogeneidad.

d) *Análisis de influencia*: El análisis de la heterogeneidad, en ocasiones, no revela ninguna causa, sino que parece ser debida a un error aleatorio, motivado porque una o varias investigaciones encuentran valores que difieren de la media. Son valores extremos sin ninguna razón aparente. Un análisis de influencia puede ser aconsejable: repetir el análisis con y sin ellos. Si el resultado significativo depende de ellos, hay que ser cauto en la inferencia. También puede cambiarse el peso de esos estudios y comprobar qué sucede.

e) *Análisis de sensibilidad:* puede hacerse de varias maneras: con un cambio en los criterios de inclusión de los estudios que se combinan (generalmente características metodológicas) y si se han utilizado criterios de corrección de errores, variando los mismos.

9. SESGO DE PUBLICACIÓN

El sesgo de publicación es el primer sesgo de selección en el metaanálisis de estudios publicados. Hay que cuantificarlo y existen varios procedimientos:

a) Representación del ln RR y su intervalo de confianza frente al tamaño de muestra del estudio: mismo comentario que en el caso anterior. Si se coloca el ln RR en el eje de ordenadas, se obtendrá la imagen de un embudo, con la parte más estrecha dirigida a la derecha (más precisión cuando el tamaño de muestra es mayor). La existencia una figura simétrica alrededor de un eje que pasa por el valor ponderado del RR habla en favor de la ausencia de este error (figura 12-2). Cuando se coloca el ln RR en el eje de abscisas se obtiene una imagen similar a un árbol de navidad. A esta imagen se le puede hacer una recta de regresión, entre el tamaño de muestra y la fuerza de la asociación (variable dependiente); es el método de Macaskill. Si la pendiente es estadísticamente significativa (distinta de cero), se considera que hay sesgo. Obsérvese en la figura 12-3 que faltan estudios pequeños con RR alrededor de la unidad; si se hace una recta de regresión la pendiente será negativa, reflejando la asimetría del gráfico (mientras que en la figura 12-2 la pendiente sería cero). Se ha comprobado que es mejor en la regresión utilizar el inverso del tamaño de muestra (procedimiento de Peters). Con un ejemplo real, en las figuras 12-4 y 12-5 se aprecia la diferencia entre los métodos de Macaskill y Peters, es mucho más sugestiva de sesgo la imagen con el inverso del tamaño de muestra.

Figura 12-2. Gráfico en embudo; riesgo relativo (RR) y tamaño de muestra, resultados hipotéticos (RR ponderado = 2). Con una recta de regresión la pendiente sería 0.

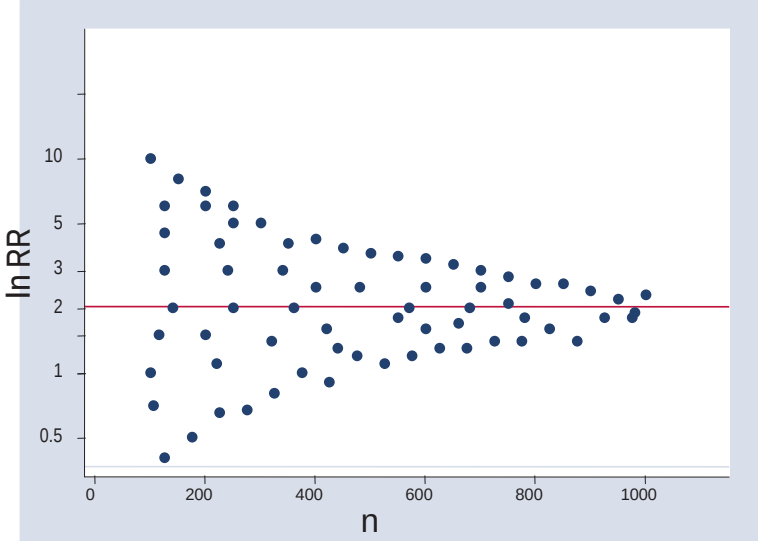


Figura 12-3. Gráfico en embudo asimétrico; riesgo relativo (RR) y tamaño de muestra,resultados hipotéticos (RR ponderado = 2). Con una recta de regresión la pendiente sería distinta de 0.

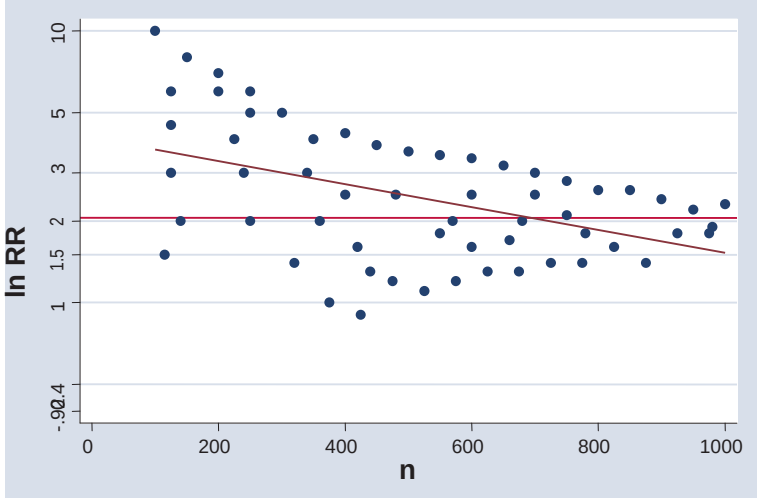


Figura 12-4. Ejemplo de presencia de sesgo de publicación por el método de regresión sobre el gráfico en embudo con n –método de Macaskill- en un metaanálisis sobre el polimorfismo de la enzima convertidora de la angiotensina y la reestenosis coronaria (Bonnici et al. BMJ 2002; 325: 517-20)

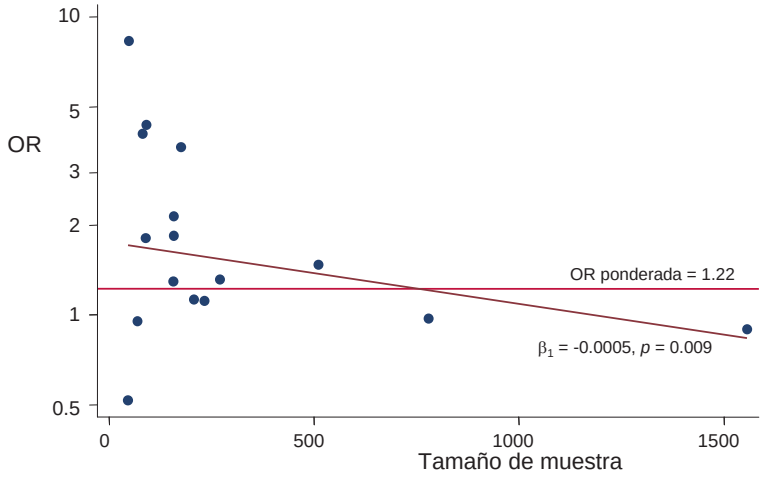
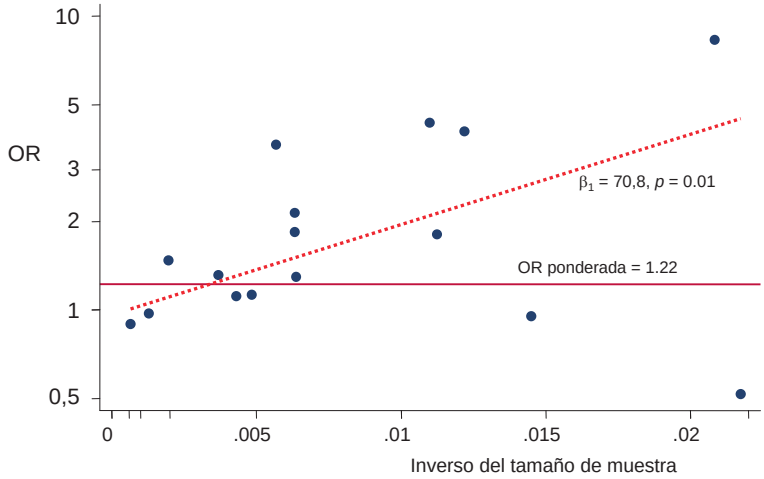


Figura 12-5. Mismo supuesto que la figura 12-4, pero utilizando 1/n como variable independiente (método de Peters)



b) Representación del logaritmo natural del RR frente a su error estándar. En este caso el embudo es más estrecho por la izquierda, cuando el error estándar es menor. Si se coloca el RR en las abscisas, en las ordenadas se toma el inverso del error estándar, para que la imagen parezca un árbol.

c) Cálculo del número probable de estudios no publicados: requieren asunciones más o menos subjetivas en su aplicación. Todos los procedimientos sólo tienen en cuenta un factor como responsable de la no publicación (lo que no es cierto): la ausencia de significación estadística en los resultados. Se puede estimar el número de estudios con resultados no significativos que se necesitarían para hacer que una asociación global significativa se hiciera no significativa estadísticamente (método de Rosenthal). También se pueden aplicar estrategias de remuestreo asumiendo que los estudios encontrados con resultado no significativo son una muestra representativa de todos los que se han hecho (publicados o no).

10. SIGNIFICACIÓN SANITARIA DEL METAANÁLISIS

La revisión sistemática (o más propiamente el metaanálisis que la precedió) ha supuesto una serie de cambios en el terreno científico. Las revisiones tradicionales incorporan cada vez más las características de sistematicidad, con consulta de un repertorio bibliográfico informatizado y con más discusión de la metodología. Si en un tema hay controversia, la revisión sistemática es una alternativa razonable para ver la contribución que tendría un nuevo estudio y las características que debiera reunir.

El metaanálisis no es una estrategia cara y sus conclusiones pueden ayudar. Contribuye a la toma de decisiones sobre la asociación entre una exposición y un efecto: (a) si en un metaanálisis, en el que al menos un estudio alcanzó significación estadística, se obtiene un resultado significativo, el metaanálisis fortalece la evidencia global; (b) si un metaanálisis, en el que ningún estudio alcanzó significación estadística, obtiene un resultado estadísticamente significativo, no es recomendable actuar según su resultado, no es prudente concluir algo significativo cuando no se observa nada claro en las unidades que la componen; (c) si en un metaanálisis en el que al menos un estudio mostró significación estadística se alcanza un resultado no significativo, habrá que reconsiderar los criterios de selección de los distintos estudios o realizar nuevos.

El metaanálisis ayuda a la enseñanza. La comprobación de que los resultados de distintas investigaciones no coinciden, apreciar que ciertas características metodológicas se asocian con el resultado, conocer que los resultados que aparecen en la literatura científica pueden estar sometidos al sesgo de publicación, etc., tiene consecuencias muy útiles para enseñar el método científico y sus limitaciones. Pero el metaanálisis tiene también errores. Las preguntas básicas que debe responder la metodología de un metaanálisis, antes de plantearse la aplicación de sus resultados, se encuentran resumidas en la tabla 12-8.

Tabla 12-8. Directrices para la valoración de revisiones de artículos publicados.

1.	¿Están claramente establecidos los objetivos y métodos?
2.	¿Se utilizaron métodos de búsqueda completos para localizar los estudios relevantes?
3.	¿Se detallaron los métodos para determinar los artículos que se incluyeron en la revisión?
4.	¿Se valoró la validez de los estudios primarios? (Análisis de la calidad)
5.	¿Fue reproducible y estuvo libre de sesgos la valoración de los estudios primarios?
6.	¿Se analizó la variación en los hallazgos de los estudios relevantes? (Análisis de la heterogeneidad)
7.	¿Se combinaron apropiadamente los hallazgos de los estudios primarios?
8.	¿Se sustentaron las conclusiones de los revisores por las evidencias citadas?

Es necesario dejar claro un hecho: el metaanálisis es tan válido como lo son los diferentes estudios que se combinan. Es imposible inferir nada si los estudios incluidos en el análisis adolecen de graves deficiencias. Aunque todos los estudios que se incluyan en un metaanálisis sean absolutamente válidos, la inferencia puede presentar problemas (aunque no haya heterogeneidad). Hay que asegurarse que los distintos estudios que se combinan pueden resumirse en un solo parámetro.

11. REFERENCIAS BIBLIOGRÁFICAS

COLLABORATION COCHRANE. www.cochrane.org

DELGADO RODRÍGUEZ M. *Revisiones Sistemáticas de Estudios. Metaanálisis.* En: JM Doménech : Diseño de Estudios Sanitarios, Barcelona: Ed. Signo SA, 2008:1-245.

EASTERBROOK PJ, BERLIN JA, GOPALAN R, MATTHEWS DR. P*ublication bias in clinical research.* Lancet 1991;337:867-872.

EGGER M, SMITH GD, ALTMAN D. *Systematic reviews in health care. Meta-analysis in context.* Londres: BMJ Books, 2001.

GREENLAND S, SALVAN A. *Bias in the one-step method for pooling study results.* Stat Med 1990;9:247-252.

GREENLAND S. *Meta-analysis.* En: KJ Rothman y S Greenland. Modern Epidemiology. 2ª ed. Boston: Lippincott-Raven, 1998: 643-673.

JÜNI P, WITSCHI A, BLOCH R, EGGER M. *The hazards of scoring the quality of clinical trials for meta-analysis.* JAMA 1999;282:1054-1060.

MACASKILL P, WALTER SD, IRWIG L. *A comparison of methods to detect publication bias in meta-analysis.* Stat Med. 2001; 20: 641-54.

MOHER D, JADAD AR, NICHOL G, PENMAN M, TUGWELL P, WALSH S. *Assessing the quality of randomized controlled trials.* Control Clin Trials 1995;16:62-73.

PETERS JL, SUTTON AJ, JONES DR, ABRAMS KR, RUSHTON L. *Comparison of two methods to detect publication bias in metaanalysis.* JAMA. 2006; 295: 676-80.

VARIOS. *Point/counterpoint: meta-analysis of observational studies.* Am J Epidemiol 1994;140:771-791.

XIII. INFERENCIA CAUSAL EN EPIDEMIOLOGÍA

José Ramón Banegas
Fernándo Rodríguez Artalejo

1. INTRODUCCIÓN

En el siglo V a.C. Hipócrates de Cos sugirió que el desarrollo de la enfermedad humana podría estar relacionado con factores de estilos de vida y con el medio ambiente externo. Es decir, enunció que las enfermedades tenían causas. Desde Sexto Empírico (200 d.C.) a Hume en 1740, muchos filósofos han reducido la cuestión de la causalidad a la de regularidad o legalidad; observamos que a un acontecimiento específico suele seguir regularmente otro, pero no observamos un nexo necesario entre ambos.

Para el filósofo Popper sí hay vínculo causal (necesario o probable) entre dos sucesos A y B: éste es la hipótesis general que postula que la relación debe darse (o darse con mayor probabilidad) en toda circunstancia, lugar y tiempo. Lo postula la hipótesis, pero no se dice que la realidad deba ser así; por eso hay que investigar desafiando la hipótesis, aunque con la esperanza de que se confirme. Así, en términos epidemiológicos, la hipótesis de que el colesterol alto (A) es causa de cardiopatía isquémica (B) no podemos justificarla racionalmente (ni lógica ni empíricamente) pues no podemos observar todos los sujetos, poblaciones o circunstancias en que podrían aparecer contraejemplos. A pesar de ello, formulamos la hipótesis (general) de relación entre A y B; que es tanto más satisfactoria, verosímil, cuanto más y mejor se vea corroborada en sucesivos y variados estudios (comprobación ayudada por ciertos criterios que luego veremos).

John Stuart Mill propuso en 1843 una serie de métodos con la pretensión de descubrir y demostrar relaciones causales. En realidad no lo logran y han sido muy criticados. Pero son métodos eliminatorios: sirven para mostrar que una determinada circunstancia no es la causa de un fenómeno dado (llamado posteriormente criterio de exclusión de factores alternativos). Y en este sentido son instrumentos para someter a ensayo una hipótesis causal; sus enunciados anticipan y describen los métodos experimentales y observacionales de la ciencia moderna.

En medicina en la segunda mitad del siglo XIX, la desmedida búsqueda de agentes infecciosos con la pretensión de ser la causa de enfermedades llevó a Koch y Henle a proponer criterios que delimitasen cuáles agentes desempeñaban realmente un papel etiológico. Este episodio lo podemos transferir de los cazadores de microbios a los actuales cazadores de factores de riesgo o de genes. No es en sí la búsqueda de nuevas causas de enfermedades lo inútil sino el hacerlo fuera de un contexto metodológico y sustantivo que le dé sentido real, práctico. Y a esto ayudan los denominados criterios de causalidad, muchas veces, pero no siempre, consustanciales e implícitos en la buena investigación científica. Con el predominio de las enfermedades crónicas en el siglo XX, los criterios de Koch y Henle dieron lugar a un desarrollo y crítica que comentamos posteriormente.

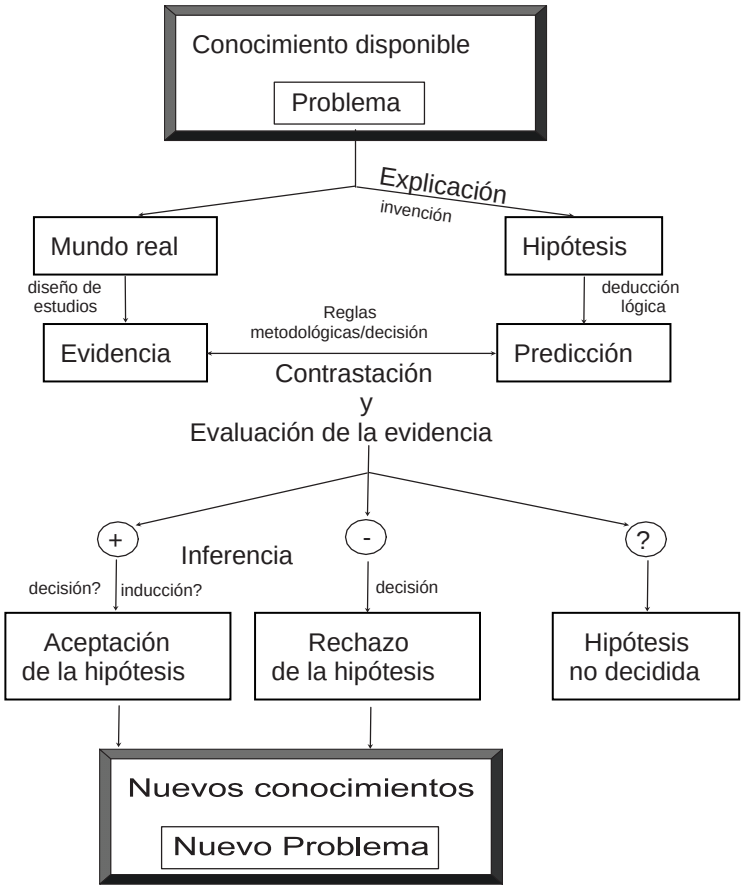
El interés por la inferencia causal en epidemiología es claro. Para MacMahon, el propósito más importante de la epidemiología es adquirir conocimiento de las causas de las enfermedades que no son prevenibles actualmente. Para ello hay que desarrollar hipótesis que expliquen la distribución de los patrones de enfermedad, y contrastar dichas hipótesis a través de estudios específicamente diseñados. De ese modo podremos clasificar a las personas enfermas en grupos que parezcan tener factores etiológicos en común.

Así pues, establecer las causas de los fenómenos de interés es uno de los principales objetivos, y problemas, de la actividad científica. Este trabajo tiene dos objetivos concretos y prácticos. El primero es responder a la pregunta: ¿qué es una causa? Para ello revisamos los modelos o representaciones formales de la causalidad de mayor interés en epidemiología y salud pública. El segundo objetivo es responder a: ¿cómo reconocemos en la práctica que una variable o fenómeno concreto es la causa de otro? Para identificar una causa se dispone de reglas metodológicas llamadas criterios de causalidad. Ninguno de ellos es suficiente para demostrar la causalidad, pero cuanto más y mejor se cumplan más verosímil será que la relación entre dos variables sea causal. Por eso en la segunda parte de este trabajo describimos los principales criterios de causalidad y presentamos un ejemplo de la aplicación de dichos criterios. Por último, ya que son muchos los criterios de causalidad existentes, concluimos con un comentario sobre la forma de combinar y organizar su aplicación en situaciones concretas.

1.1. Aspectos filosóficos

Cuando se aborda el tema de la causalidad se suele titular y comenzar hablando de inferencia. Sin embargo, la inferencia (causal) podría comprenderse mejor si la ubicamos en el marco general, más amplio y más adecuado de la explicación (causal) (figura 13-1). Explicación, contrastación e inferencia siempre entran en juego en el método científico, aunque dependiendo de cuál es nuestra situación problemática: la búsqueda de una hipótesis (causal) explicativa, la contrastación de una hipótesis ya formulada, o la inferencia sobre una hipótesis contrastada. Pero la inferencia siempre es precedida por la explicación y consiste esencialmente en valorar los resultados de las contrastaciones.

Figura 13-1. Esquema general del método científico. Modificado de Popper, Bunge y Giere.



Fte: Elaboración propia a partir de Popper, Bunge y Giere.

La inferencia para algunos autores sólo hace referencia a la lógica. Pero la llamada inferencia causal, el paso de la evidencia a la hipótesis, no parece ser un proceso lógico. Es una valoración de la hipótesis basada en observaciones, experimentos y argumentos relevantes. Y para la valoración de esas evidencias se ayuda de reglas metodológicas, es decir de elementos no lógicos que ayudan a la decisión sobre si la evidencia apoya o no a la hipótesis causal, y en qué medida. Una decisión no es un elemento lógico, más no por eso es arbitrario; se toma sopesando motivos, razonadamente; disminuyendo el riesgo de error y eliminando la subjetividad de la decisión en la discusión con otros científicos. La evidencia (y por tanto la hipótesis) se decide, pues, por un acuerdo o convención razonado, reglamentado y objetivo. En caso de duda hay que renunciar a tomar un acuerdo y proseguir la contrastación y discusión.

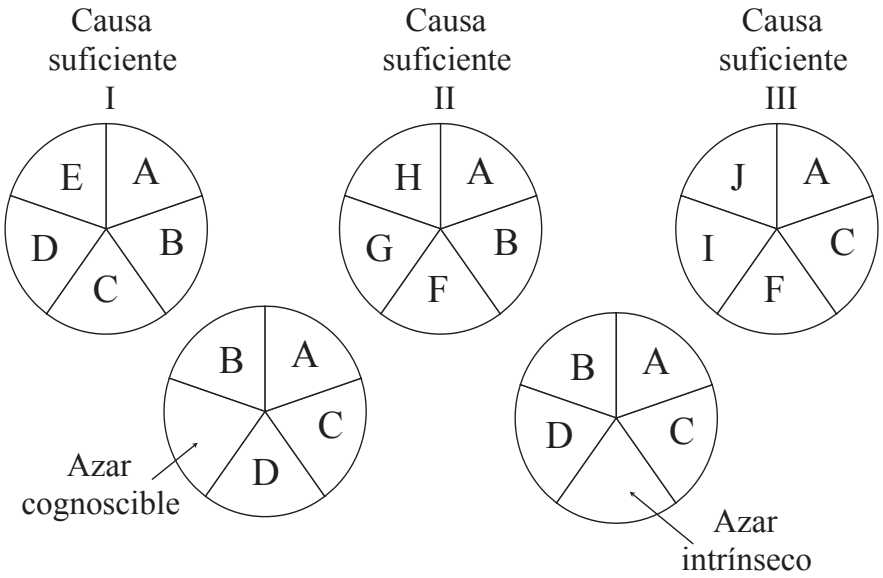
Precisamente, que sean decisiones y no conclusiones (lógicas) lo que en último término decide el destino de una hipótesis (causal), habla de que la evaluación de la hipótesis es tentativa, provisional, no concluyente o definitiva, tal como decía Popper.

2. MODELOS CAUSALES EN EPIDEMIOLOGÍA

2.1. El modelo determinista modificado

Algunos autores, como Rothman, propugnan un modelo general de causación que permite relacionar algunos de los principales principios epidemiológicos: fuerza de los efectos, interacción entre causas, proporciones de la enfermedad debidas a causas específicas, periodo de inducción, etc. El modelo propuesto es un modelo teorético, es decir imaginario, pero que ayuda a visualizar cómo pudieran ser las cosas y, por tanto, alienta a comprobarlas. En este modelo, causa se define como el suceso, condicion o característica, que inicia o permite, sola o en conjunto con otras causas, una secuencia de sucesos que resulta en un efecto. Se trata de un modelo determinista modificado. Es determinista porque supone que una enfermedad puede tener una o más causas suficientes (CS), cada una de las cuales, por sí mismas, determina inexorablemente la enfermedad en un individuo. Cada CS (o mecanismo causal teórico) está compuesta de una constelación mínima de causas componentes (CC) que interactúan para producir un efecto (figura 13-2). Mínima significa que cada una de las CC es necesaria para producir la enfermedad. El problema es que en el estado actual del conocimiento, para la mayoría, si no todas, las enfermedades conocemos solo algunos de sus componentes. Por eso el modelo atempera el determinismo modificándolo, dando cabida al azar en forma de factores o CC todavía no identificados. No conocemos una CS ni siquiera del cáncer de pulmón. Así, que un individuo fumador, o expuesto a radón o asbestos, desarrolle o no cáncer de pulmón, podría deberse a que actúen o no otros factores causales todavía desconocidos, necesarios para completar una CS. Por último, se denomina causa necesaria a aquella CC que está presente en todas las CS de una enfermedad , y sin la cual, por tanto, no se produce ningún caso de enfermedad.

Figura 13-2. Modelo causal determinista modificado. Modificado de Rothman, 1998.



Fte: Rothman y elaboración propia.

Por el contrario, un modelo indeterminista diría que no solo hay azar por ignorancia, debido a factores desconocidos, sino que el azar es intrínseco a la situación, está abierto a múltiples y no predeterminadas circunstancias (figura 13-2). Dado que para cada enfermedad hay múltiples factores implicados, con múltiples interacciones posibles, no podemos predecir qué ocurrirá en un caso concreto. El futuro está indeterminado, abierto. No es solo un problema de ignorancia de los factores implicados.

El modelo es útil para orientar la investigación epidemiológica y la toma de decisiones. Según este modelo, un factor es una causa de un evento si su alteración conlleva una variación de la frecuencia de la enfermedad, pues desde un punto de vista pragmático la epidemiología persigue descubrir relaciones que ofrezcan posibilidades para la prevención. La causación necesaria y suficiente son simplemente extremos de esta definición. En el momento actual, la mayor parte de los factores causales conocidos no son necesarios ni suficientes para producir la enfermedad: la actuación del factor causal incrementa la frecuencia del resultado; pero el resultado no siempre ocurre, y puede ocurrir sin la actuación del factor. Por eso muchos se inclinan a hablar solo de factores de riesgo (causales).

Este modelo acomoda también algunos principios epidemiológicos (figura 13-2):

1. No es necesario conocer todas las CC para prevenir la enfermedad. Eliminando uno de los elementos de la CS se previenen todos los casos de enfermedad (o fraccion etiológica, FE) que esta CS origina. Así, si una CS de una enfermedad tiene cinco CC, a saber, A, B, C, D, E; y la FE de la CC "D" es 20%, la FE de la/las CS de la que forma parte es también 20% pues sin "D" la CS no se completa y, por tanto, la enfermedad no se produce. Eliminando "D" se eliminan el 20% de los casos de esa enfermedad. De forma similar, la eliminación del uso de jeringuillas contaminadas (solo una causa componente) prevendría la aparición de un gran número de casos de Sida, hepatitis viral, etc.
2. La suma de las FE de todas las CS de una enfermedad es siempre el 100%. Sin embargo, la FE correspondiente a cada CC equivale a la suma de las FE de la/las CS de las que forma parte. Así, si una enfermedad tiene tres CS, siendo la FE de la CS primera FE (I) = 20%, y la de la tercera FE (III) = 50%, entonces la FE (II) = 30%. Y si la CS (I) la forman las CC A, B, C, D, E; la II las A, B, F, G, H; y la III, las A, C, F, I, J, entonces,

FE (D) = FE (E) = FE (I) = 20%

FE (B) = FE (I) + FE (II) = 50%

FE (A) = FE (I) + FE (II) + FE (III) = 100%

FE (A) + FE (B) + FE (C) + ... + FE (J) = 500%

Parece colegirse que en salud pública debería considerarse en la priorización, aquellas CC que por estar implicadas en varias CS responderían de una mayor proporción de enfermedad.

Que la suma de las FE de todas las CC pueda superar el 100% no significa obviamente que pueda prevenirse más del 100% de una enfermedad. No hay límite superior a la suma de las FE. Solo la FE atribuible a una causa aislada no puede exceder el 100%. Y solo una causa necesaria contribuirá al 100%. Por tanto, para cualquier enfermedad, la suma de las FE de las CC es al menos del 100%. Así, se ha declarado que los porcentajes de casos de cáncer atribuibles a factores de riesgo es: a la alimentación insalubre el 10-70%, al tabaco el 25-35%, a los tóxicos laborales el 10-40%, a los tóxicos ambientales el 5-10% y al alcohol el 10-20%. O el porcentaje de cardiopatías atribuibles a factores de riesgo es: al colesterol sérico el 30-40%, al tabaco el 20-25%, a la hipertensión el 20-25%, y a la inactividad física el 10-20%.

3. La medida en que una CC afecta a la frecuencia de la enfermedad depende de la frecuencia relativa de las restantes CC pertenecientes a la misma CS en la población. Así, si en la CS I de una enfermedad, la CC "E" está presente en el 50% de la población, y el resto de CC (A, B, C, D) están presentes de forma combinada en la población en un 20%, entonces "E" contribuirá con 0,50 x 0,20 = 10% de casos de enfermedad en la poblacion. Si, por el contrario, "E" sigue presente en el 50% de la población, y el resto de CC de forma combinada en el 2% de ella, entonces la presencia de "E" conduciría a la enfermedad en un 1% de los individuos. Del mismo modo, la fuerza o riesgo relativo (RR) de una CC depende de la frecuencia relativa de las restantes CC pertenecientes a la misma CS en la población.
4. Existe interacción entre las CC de una CS, que depende de la prevalencia relativa de dichas CC en la población. Se habla de sinergismo si el efecto conjunto es superior a la suma de los efectos por separado. Y de antagonismo si es al revés. Es decir, hay interacción (bajo un modelo aditivo) si el riesgo atribuible (RA) a la exposición conjunta de dos CC (A y B) es superior a la suma de sus RA por separado: RA (A+B) > RA (A) + RA (B). Y antagonismo, si es al revés.
5. Los periodos de inducción no son constantes para las enfermedades, sino específicos de cada CC. Las CC actuando al inicio de la secuencia causal tienen un periodo de inducción largo. Las CC finales, breve. Además, no hay tiempos de inducción constantes para una enfermedad en relación con sus CC. Dosis pequeñas requieren un conjunto de CC para completar una CS mayor que las dosis grandes. Éstas tienen un periodo de inducción más corto al precisar menos CC para completar una CS.

2.2. Otros modelos causales. Modelo de "maraña" o red

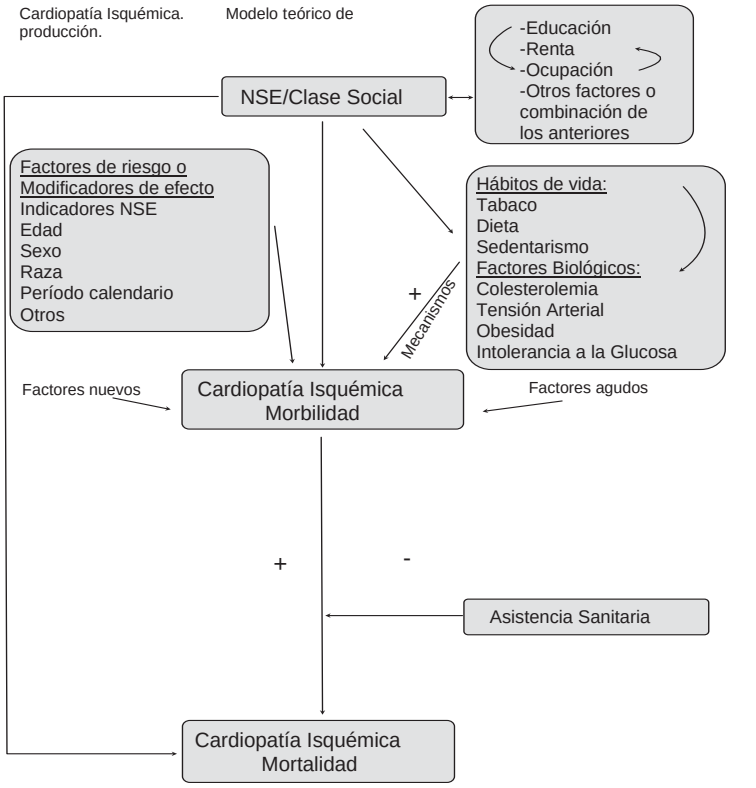
A pesar de las bondades del modelo anterior, se dispone actualmente de escasa información sobre las causas de muchas enfermedades. De ahí el interés de otros modelos causales más clásicos. En epidemiología, muchos de los factores causales que se estudian no son mecanismos bioquímicos o moleculares de las enfermedades, aunque éstos brinden plausibilidad a las relaciones epidemiológicas. Sin embargo, dichos mecanismos son también causas. El modelo

multicausal de la "maraña" o red acomoda de forma secuencial las causas distales y proximales de la enfermedad en un solo esquema.

Por lo general una enfermedad tiene varias causas que, conjunta o independientemente, dan lugar a la aparición de la misma. De igual manera una sola causa puede contribuir a la aparición de varias enfermedades diferentes. Las causas de las enfermedades pueden dar origen a éstas de forma indirecta, es decir, a través de la producción de otras causas de enfermedad, o directamente. En realidad casi siempre hay indicios de causas intermediarias; por eso la relación de causalidad está mejor caracterizada por el grado de proximidad que por el carácter directo o indirecto de la misma. De hecho, el que una asociación causal sea directa o no depende del estado actual del conocimiento de los mecanismos de una asociación particular, por lo que no tiene implicaciones necesarias para la prevención. La llamada "red causal" es un modelo que ilustra de forma secuencial las relaciones entre causas y enfermedad. La figura 13-3 muestra simplificadamente la red causal o modelo teórico de producción de la cardiopatía isquémica. Las causas que ocupan en la red los niveles jerárquicos más altos, a menudo características sociales, determinantes a su vez de estilos de vida o factores relacionados, son del máximo interés desde el punto de vista epidemiológico y preventivo. El modelo teórico es un modelo reduccionista (pueden existir varias redes causales para una enfermedad dada) que ayuda sobre todo a identificar los elementos implicados (principales, confusores, modificadores de efecto, etc.) en estudios concretos, y ayuda a comprender que la prevención (reducción) de la enfermedad depende de la eliminación de un elemento troncal de la red.

Este modelo está relacionado con el modelo determinista modificado. Cada uno de los eslabones en la cadena causal pueden ser componentes de causas suficientes. Pero cada una de las causas componentes tiene sus propias redes causales; y una enfermedad determinada puede resultar de varias redes de causas suficientes, que pueden tener o no componentes en común.

Figura 13-3. Modelo causal de "maraña" o red.
Secuencia de producción de la cardiopatía isquémica. Elaboración propia.



NSE: nivel socioeconómico. Fte: Elaboración propia.

3. LOS CRITERIOS DE CAUSALIDAD

3.1. Los criterios de Henle-Koch

Cuando las principales enfermedades eran infecciosas, la cuestión se centraba en saber qué evidencia era necesaria para probar que un organismo causaba la enfermedad. Henle propuso en 1840 postulados de causación, expandidos por Koch en los 1880s (tabla 13-1). Sin embargo, conforme las enfermedades no infecciosas ganaron importancia a mediados del siglo XX, se replantearon estos criterios. En 1954 E. Cuyler Hammond enunció una serie de criterios, posteriormente discutidos y ampliados por Yerushalmy y Palmer en 1959, Sartwell y otros. En 1964 un comité de expertos nombrado por el "Surgeon General" de Estados Unidos de América desarrolló directrices para evaluar la evidencia sobre la relación entre tabaco y cáncer de pulmón. Estas guías han sido revisadas desde entonces. Por ser los más conocidos y usados en epidemiología, comentamos de forma detallada los criterios que Sir Arthur Bradford Hill publicó en 1965, para valorar la posible causalidad de una asociación.

Tabla 13-1. Criterios de causalidad según diversos autores

Hume, 1739
1. Contigüidad en tiempo y lugar
2. Prioridad en el tiempo
3. Conjunción constante

Henle-Koch, 1882
1. Agente aislado en cultivo puro en cada caso de enfermedad
2. Ausencia del agente en otras enfermedades
3. Reproducción experimental animal de la enfermedad
4. Agente recuperado del caso experimental producido

MacMahon, 1960
1. Secuencia temporal
2. Fuerza de la asociación
3. Consistencia con conocimiento previo

Surgeon General de EE.UU., 1964
1. Consistencia
2. Fuerza
3. Especificidad
4. Relación temporal
5. Coherencia

Pooling Project, 1978
1. Consistencia
2. Fuerza
3. Gradualidad
4. Independencia
5. Relación temporal
6. Generalizabilidad predictiva
7. Coherencia

Kleinbaum et al, 1982 (factor de riesgo causal)
1. Debe covariar con la enfermedad (asociación estadísticamente significativa)
2. Debe preceder en el tiempo a la enfermedad
3. Asociación no debida enteramente a una fuente de error (sesgos o confusores)

3.2. Los criterios de Hill

Los criterios o estándares se dirigen a la evaluación de la evidencia disponible sobre la relación entre dos factores, típicamente la exposición a un agente y una enfermedad. No comentaremos los criterios en el orden en que Hill los enumeró (tabla 13-2); aunque los enumeró, Hill no estableció una jerarquía entre ellos. Según dijo, "ninguno de los nueve puntos de vista es incuestionable para aceptar o rechazar la causación". Además, los comentamos más allá de lo descrito por Hill, haciendo referencias a su moderna valoración. Como se aprecia en la tabla 13-1, muchos de los criterios se corresponden con los enunciados antes y después por otros autores.

Aunque muchos autores (y el propio Hill) recomiendan usar los criterios causales solo ante una relación estadísticamente significativa, matizan esta cuestión. Por ejemplo, si el tamaño muestral es pequeño hay que interpretar una diferencia sustantiva no significativa como "no probada" y no como "no real o no causal". Además, Hill resaltaba que los criterios causales solo eran necesarios ante evidencia observacional. Ante buenos datos experimentales, en un ensayo bien realizado, la existencia de significación estadística bastaría para inferir una relación como causal. Pero, al margen de las dificultades prácticas y éticas, muchos experimentos reales, no ideales, no garantizan automáticamente que las variables extrañas hayan sido controladas. El propio Hill matizó esto en algunos de sus escritos. Los criterios, al menos algunos (por ejemplo, consistencia o plausibilidad), pueden seguir siendo útiles. En definitiva, dice Hill, cuando la diferencia observada es muy grande o muy constante, sería absurdo atribuirla al azar y carece de importancia un test de significación; y si es muy pequeña no interesa si técnicamente es significativa o no. Pero entre estos dos extremos, los tests ayudan en la interpretación.

Tabla 13-2. Criterios de causalidad de A.B. Hill, 1965*

Criterio	Descripción
1. Fuerza	Riesgo relativo grande
2. Consistencia	Asociación observada repetidamente por varias personas, en sitios, circunstancias y épocas diferentes
3. Especificidad	Una causa lleva a un solo efecto
4. Temporalidad	La causa precede al efecto
5. Gradiente biológico	La magnitud de la enfermedad aumenta con la magnitud de la exposición a la causa
6. Plausibilidad	La asociación tiene sentido de acuerdo al conocimiento biológico del momento
7. Coherencia	Ausencia de conflicto con la historia natural y biológica de la enfermedad
8. Experimento	La reducción de la exposición a la causa se asocia a una disminución de la enfermedad
9. Analogía	Relación causa-efecto ya establecida para un agente-enfermedad similares

** Criterios enumerados en el orden propuesto por Hill.*

3.2.1. Temporalidad

La exposición al factor presuntamente causal debe preceder en el tiempo a la aparición del efecto. Pues podría ocurrir que fuera la consecuencia del "efecto"; así, muchos estilos de vida (dieta, por ejemplo) cambian tras (y a consecuencia de) las etapas iniciales de la enfermedad que presuntamente causan, sobre todo si es una enfermedad crónica de instalación lenta. Sin embargo, que la "causa" siga al efecto no significa que la "causa" no pueda causar ese efecto en los casos en que le precede (la dieta desequilibrada puede causar enfermedad aunque en una situación dada se altere tras los primeros síntomas de la enfermedad asociada).

Un refinamiento útil de este criterio es considerar que el tiempo en que la causa putativa precede al efecto es relevante, es decir compatible con el periodo de inducción de la enfermedad o consistente con un mecanismo biológico. No todos los estudios epidemiológicos tienen la misma capacidad para poner de manifiesto este criterio. Por ejemplo, si el estudio es de cohorte prospectivo el criterio es más fácilmente comprobable que si es de casos-control retrospectivo. Si el estudio es transversal, el criterio queda garantizado solo si el factor de exposición es una característica fija (por ejemplo, el grupo sanguíneo o la fecha de nacimiento). No solo el tipo de diseño, también la medida de frecuencia elegida tiene diferente eficacia para demostrar la secuencia cronológica: es mejor la incidencia que la prevalencia, y la prevalencia de punto a la de sobrevida.

Este es el único criterio considerado por algunos autores como *sine qua non*. Sin embargo, el propio Hill no lo consideró imprescindible, probablemente por la dificultad de su consecución en ocasiones. Ser taxativo en la exigencia de este criterio llevaría a despreciar muchos estudios observacionales. Además, el periodo de inducción de la última causa componente de una causa suficiente cuyos componentes actúen secuencialmente (algunos cánceres) es prácticamente cero, es decir la causa sería prácticamente simultánea con el efecto. Por otro lado, el criterio es muy obvio con algunos tipos de estudios (experimentos) o con variables fijas.

3.2.2. Fuerza

También denominado magnitud o intensidad de la asociación o tamaño del efecto, es el primer criterio que Hill puso en su lista, y es el grado de estrechez de la asociación entre el factor y la enfermedad. En epidemiología, la fuerza de la asociación es usualmente medida por el grado en que el RR o el odds ratio se separan de la unidad, sea por encima de 1 (en el caso de exposiciones que causan enfermedad), o por debajo de 1 (en el caso de intervenciones preventivas). Pero también puede estimarse por una diferencia de riesgos, de medias o de proporciones, o por un coeficiente de regresión. La medida relativa del riesgo parece preferible a la diferencia absoluta para valorar la fuerza de la asociación por ser más informativa: los números relativos son más elocuentes en relación con la etiología. Así, en la relación del tabaco con el cáncer de pulmón, un RR de 8 (ratio de tasas 0,57/0,07 x1000 y año en fumadores de 1-14 cigarrillos/día frente a no fumadores) es más elocuente que una diferencia de tasas de 0,50x1000 y año. Asimismo, una diferencia de 2,2 a 2,6 por mil de mortalidad por cáncer de pulmón y cardiopatías entre dos grupos (fumadores y no fumadores) es menos probable que sea debida a un error de medición si se presenta entre las tasas 2,27 y 0,07 (cáncer de pulmón, RR=32) que entre las tasas 9,9 y 7,3 (enfermedad cardiovascular, RR=1,4).

La regla de interpretación de este criterio es que cuanto mayor es la fuerza más verosímil es la hipótesis causal (o más "probablemente" causal es, se suele decir). Un RR muy alto (por ejemplo, superior a 10 como para el tabaco y el cáncer de pulmón) solo puede deberse a una variable de confusión si es mucho más frecuente en los expuestos que en lo no expuestos y se asocia fuertemente con la enfermedad. Los confusores rara vez dan cuenta de una diferencia tan considerable. Además, para Hill, un RR muy alto requiere un factor tan íntimamente

ligado al factor estudiado que sería fácilmente detectable. Inversamente, RR próximos o inferiores a 2 (como en la relación tabaco-trombosis coronaria) inducen a pensar en factores de confusión (FC).

Se requiere experiencia en el área de investigación para apreciar cuándo un RR u otra medida del tamaño del efecto es realmente grande, importante para constituir buena evidencia de causalidad. A esta valoración puede ayudar el examinar cuán consistentemente grande es el efecto a lo largo de una serie de estudios. En cualquier caso, autores como Brownson o Hill hablan, en el contexto de las principales enfermedades crónicas, de RR débiles si son inferiores a 2; moderados entre 2 y 4, y fuertes si son superiores a 4 (estos RR suelen estar ajustados para varios confusores). Weed ha propuesto recientemente, en el campo de la epidemiología nutricional, que RR estadísticamente significativos por encima de 1,2 o por debajo de 0,8 son valorables en el juicio causal, dada la rareza de encontrar RR mayores de 1,2 o menores de 0,8; un RR de 1,4-1,5 (o 0,5-0,6) sería fuerte. Otros autores, por el contrario, son más exigentes y solicitan RR de al menos 4 o 5 para empezar a considerar que una relación sea causal. Sin embargo, exigir valores concretos al RR equivale a olvidar que los criterios no son reglas lógicas, autosuficientes. La valoración del tamaño del efecto es una cuestión de juicio informado.

Como todos los criterios, el de fuerza puede tener contraejemplos, excepciones. El propio Hill reconoce que una asociación débil puede ser causal. Así lo es la relativamente débil relación entre consumo de cigarrillos y enfermedad cardiovascular (hay otras muchas pruebas de ello); otro caso similar es el del tabaquismo pasivo y cáncer de pulmón. También una relación fuerte puede ser no causal, debido a FC no controlados. Es el caso del síndrome de Down y el orden de nacimiento confundido por la edad materna. Por supuesto, una vez que el confusor está identificado, su control reduce la asociación.

Otra limitación del RR (u otra medida de la intensidad de la asociación) es que su magnitud depende de la prevalencia de las otras causas componentes necesarias para completar la causa suficiente o mecanismo causal del que forma parte. Cuanto más frecuentes son las causas complementarias de un factor, mayor es la fuerza de éste y al contrario. Así, el RR es diferente entre poblaciones con distinta distribución de factores de riesgo: el RR de enfermedad coronaria por hipercolesterolemia en dos poblaciones con similar frecuencia de la misma, será mayor en aquella que tenga mayor proporción de fumadores o hipertensos (las tres variables forman parte de la misma causa suficiente de la enfermedad, digamos la placa de ateroma).

En conclusión, una asociación fuerte sólo sirve (ni más, ni menos) para descartar la hipótesis de que la asociación es debida enteramente a un confusor u otra fuente de sesgo modesto. Pero cuanto más específico es el constituyente de una causa, más cerca del evento biológico está, o más específica o precisa es la medida de la exposición, más fuerte es la asociación. Cuando, como es frecuente, la variable de exposición medida es una estimación muy indirecta del verdadero factor biológico de interés, la asociación es débil. Por ejemplo, medir un alimento en lugar de una vitamina. Cuando el factor evaluado es un estimador más próximo del agente biológico verdadero, el RR será mucho más alto. Esta es otra razón por la que un RR fuerte es más verosímilmente causal.

3.2.3. Consistencia

Para algunos es el criterio más importante de causalidad. No basta un resultado aislado para refutar o corroborar en la práctica una hipótesis (causal). Es decir, aunque la evidencia circunstancial puede “apoyar” positiva o negativamente una hipótesis, ni el mejor y mayor

estudio aislado descarta verosímilmente el error; los estudios aislados son raramente definitivos.

En epidemiología no es posible el grado de control conseguible en el laboratorio y, por tanto, el sustituto de la replicación exacta es la repetición en condiciones similares: basta que se reproduzcan razonablemente las condiciones relevantes del estudio (sujetos comparables, misma variable de resultado, etc.) para comprobar si el efecto es similar, es decir, si hay consistencia. Hallazgos similares obtenidos en diferentes poblaciones, bajo circunstancias y con métodos distintos, y por diferentes investigadores, tienen un peso mayor para apoyar la causalidad. Es poco probable que estudios con errores diferentes concluyan una misma cosa si ésta no es cierta. Por ejemplo, el juicio causal sobre la relación entre el consumo actual de cigarrillos y el riesgo de enfermedad coronaria se fortaleció cuando se realizaron un gran número de estudios casos-control y de cohortes en los últimos 40 años en un amplio rango de poblaciones de culturas distintas, observando a millones de personas, y todos ellos mostraron consistentemente un riesgo aumentado.

Hay dos tipos de replicación. La primera es la mera repetición, con ánimo de verificar apresurada o fuertemente un resultado. La segunda es la que varía las circunstancias o métodos para ver si la hipótesis resiste nuevas pruebas independientes, duras, severas (*ordeals*): por ejemplo, con medición más precisa (técnicas más fiables) de la exposición o de la enfermedad, poblaciones de distintas características sociodemográficas, o consideración de otros factores "causales" o FC, etc. Esta segunda es la más rica y útil para la inferencia causal. Por eso la contrastación de la hipótesis de Barker sobre los orígenes fetales de la enfermedad coronaria ha sido acusada de *verificacionismo* -lo que no quiere decir que no sea cierta-: para algunos sólo se han buscado pruebas a su favor, pero no se ha contrastado duramente, para ver si surgen contraejemplos (controlando el nivel socioeconómico a lo largo de la vida, o estudiando gemelos; o buscando mecanismos plausibles). En términos filosóficos, la refutación es más fuerte que la verificación. Las pruebas afirmativas son más fáciles de buscar y encontrar (sobre todo en los primeros estudios); y mientras que un conjunto finito de observaciones, si son verdaderas y contradicen a la hipótesis, puede refutarla, bajo ninguna condición podría verificarla al concordar con la misma (siempre habrá casos inobservados que pueden contradecirla).

La cuestión de cuándo una relación es consistente es materia de juicio aunque, como en otros criterios, no es un asunto arbitrario. Además, que no haya un criterio nítido no significa que no haya una idea reguladora, una regla que nos oriente y que facilite el acuerdo. Una regla razonable puede ser: para ser consistentes, los hallazgos obtenidos en estudios comparables y correctamente realizados deben ir en la misma dirección (RR por encima o por debajo de 1). Aunque los resultados pueden diferir en tamaño, y esto puede tener trascendencia práctica, ello va más allá de la valoración de la causalidad. Sin embargo, nadie duda que es deseable que los hallazgos "consistentes" tengan una magnitud (por ejemplo, RR) similar.

Este criterio tiene también sus excepciones. Si no se repite la asociación, no puede descartarse una relación causal porque algunos efectos ocurren solo bajo circunstancias inusuales. Más precisamente, el efecto de un agente causal no puede producirse a menos que actúen antes o después las causas componentes, que se complementan hasta producir una causa suficiente. Estas condiciones no son siempre satisfechas. Así, el contacto con el virus HIV puede causar infección, pero no siempre lo hace: el virus tiene que estar presente en una concentración determinada y en una zona permeable a su entrada (zona de la piel con pérdida de integridad por ejemplo). Además, en ocasiones ocurren acontecimientos singulares, como el exceso de cáncer de pulmón y senos nasales en trabajadores de refinería de níquel en Gales del Sur a principios de siglo. Aparte de esto, son frecuentes los resultados inconsistentes al repetir estudios. No se trata de comparar el número de estudios con resultado positivo con el de los

negativos. Cada estudio debe ser evaluado según su calidad, sesgos y otros criterios que veremos; y hay que ponderarlos, siquiera sea por juicio cualitativo. Los resultados negativos que sean falsos pueden ocurrir por tamaño muestral pequeño, sesgos o FC. Un estudio que falle en encontrar una asociación, pero que es limitado en metodología y tamaño, de modo que no tiene poder suficiente para detectar un efecto (si existe), es de poco valor; por eso es de dudoso valor que un resultado nuevo y controvertido sea contradicho por estudios débiles, mal diseñados o pequeños. A veces, los resultados aparentemente contradictorios pueden explicarse por la presencia de modificación de efecto, es decir, porque la exposición tiene efectos distintos en diversos sectores de la población. En cualquier caso, si los hallazgos de varios estudios son inconsistentes, la inconsistencia tiene que ser explicada. Por último, si la asociación se repite podría deberse a un sesgo siempre en la misma dirección. Por tanto, ni la ausencia ni la presencia de consistencia garantiza la causalidad.

Por último, la consistencia está muy relacionada con la generalización (validez externa o aplicabilidad de los resultados a otras poblaciones). La repetición en diversas circunstancias es el mejor test de generalizabilidad; y la consistencia, de generalización. Por ejemplo, el estudio de Framingham para examinar las causas de la enfermedad coronaria se diseñó originalmente en individuos de edad media, de clase media, en una muestra sistemática en la pequeña ciudad norteamericana de Framingham. Pero sus resultados (p.e., la hipercolesterolemia como factor de riesgo cardiovascular) se aplican en otras partes del mundo. Y ello no solo porque la hipótesis es universal por la constancia de las relaciones biológicas subyacentes (y, por ende, de los patrones de enfermedad y RR) entre diferentes individuos, sino también porque se han obtenido resultados similares al volver a examinar esas relaciones o hipótesis en otras poblaciones, otras circunstancias y con otros métodos de estudio. Además, la generalización no es solo una manera de examinar la aplicabilidad de los resultados sino, además, un método de enriquecer o elaborar la hipótesis causal, de corroborarla más fuertemente, pues una hipótesis más general ha probado su temple en más circunstancias.

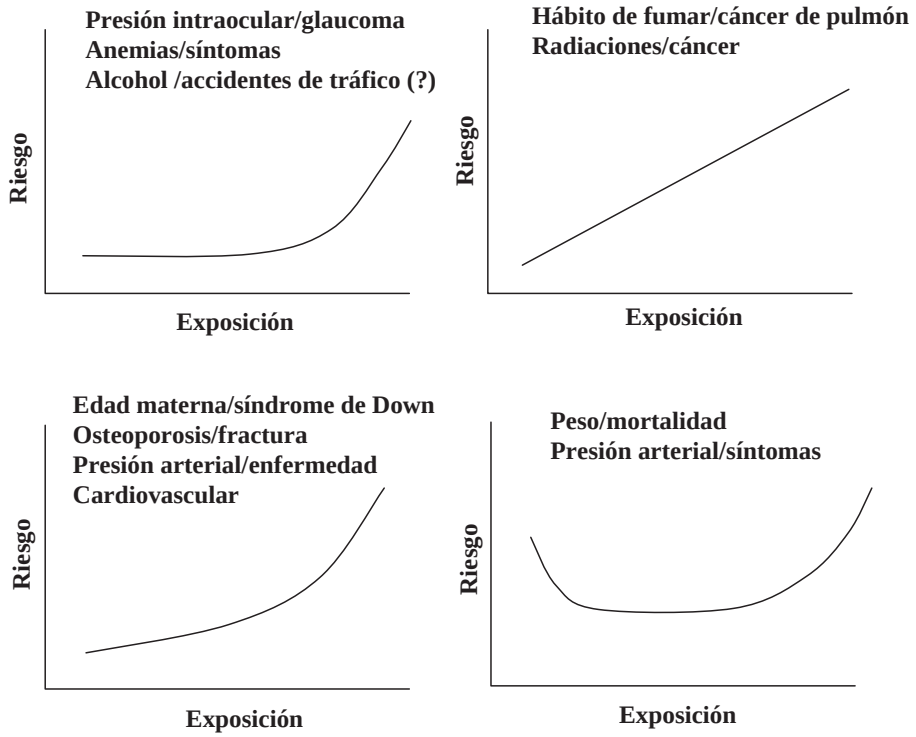
3.2.4. Gradiente biológico

También llamado relación dosis-respuesta. Para algunos autores es una variante del criterio de magnitud. Nos dice que si un factor es causal, cuanto mayor es la exposición al factor mayor es el riesgo de enfermedad. Es decir, hay una relación monotónica entre el factor y su efecto.

Si es posible, las variables predictoras deben medirse de forma continua o mediante varias categorías, para estudiar la relación dosis-respuesta. La asociación entre fumar cigarrillos y cáncer de pulmón es un ejemplo de esta relación: los fumadores moderados tienen tasas de cáncer más elevadas que los no fumadores, y los grandes fumadores presentan tasas aún mayores. También la relación de las radiaciones con el cáncer parece ser aproximadamente lineal a lo largo de todo el intervalo de exposición. Sin embargo, otras asociaciones producen relación curvilineal (p.e. osteoporosis/fractura, presión arterial/enfermedad cardiovascular), y otras son en forma de U o J (p.e. peso/mortalidad) (figura 13-4).

Hay relaciones causales que no cumplen este criterio. Algunos efectos solo se producen por encima de un umbral de exposición, como en el caso de las relaciones presión intraocular/glaucoma, o dietilestilboestrol y cáncer de vagina. Tampoco las asociaciones que muestren relación dosis-respuesta son necesariamente causales, pues puede solo reflejar el efecto de un confusor no controlado. Por ejemplo, la relación dosis-respuesta entre el tabaco y la cirrosis hepática se explica por la fuerte asociación monotónica entre el tabaco y el consumo de alcohol.

Figura 13-4. Relación dosis-respuesta en cuatro situaciones. Tomado de Rose, 1992 (La estrategia de la medicina preventiva. Masson, 1994).



Relación dosis-respuesta en cuatro situaciones. Fte: Rose G. La estrategia de la medicina preventiva. Masson, 1994.

3.2.5. Plausibilidad

El apoyo que se "exige" a una hipótesis causal no tiene por qué consistir solamente en datos que corroboren sus implicaciones (convalidación empírica); puede venir también de hipótesis o teorías más amplias que tengan apoyo empírico independiente. Si se añade esta convalidación teórica, la explicación es más profunda, más satisfactoria. En una palabra, no nos conformamos con correlaciones consistentes, universales, explicativas de hechos o problemas, queremos saber el mecanismo que las explica.

Un mecanismo biológicamente plausible, conocido o postulado, debería poder explicar por qué la exposición altera el riesgo de desarrollar la enfermedad. Sin embargo, dado que su consideración depende del estado actual del conocimiento, su falta no descarta la causalidad. Un ejemplo es la asociación entre el tabaquismo y el cáncer de cervix. Al principio le faltaba verosimilitud biológica y se creyó poco en ella, pero luego se identificaron componentes del humo del tabaco en el moco cervical y en otros líquidos corporales de las fumadoras. De hecho, las observación aparentemente implausible alentó la búsqueda de mecanismos. Por otra parte, su presencia no garantiza causalidad. Un mecanismo aparentemente plausible para la asociación puede luego no demostrarse, y la asociación puede estar sesgada. Además, una hipótesis epidemiológica puede ser compatible con varias teorías mecanísmicas.

Una teoría mecanísmica es preferible a una que no considera el mecanismo, por ser lógicamente más fuerte (implica a ésta), más rica (mayor contenido), metodológicamente más exigente (da más numerosas y precisas oportunidades de contrastación), y suministrar una explicación más profunda (detalla un nivel de la realidad distinto, más básico). Un ejemplo es la teoría oxidativa de las LDL como mecanismo para explicar la relación entre colesterol sérico y cardiopatía isquémica.

La ausencia de este criterio tiene que ver con las llamadas relaciones o hipótesis de "caja negra", que ignoran los mecanismos de la relación observada, y que algunos identifican con la epidemiología de factores de riesgo. Las relaciones de caja negra pueden ser valiosas; hay muchos ejemplos de ello. Además, son más generales, sencillas, precisas y seguras, por lo que muchos investigadores siguen proponiéndolas. Pero buscar su mecanismo puede lograr un mejor conocimiento del problema y, por tanto, mejorar sus posibilidades de control. Aunque Snow no conocía la implicación de un bacilo en la epidemia de cólera de Londres el pasado siglo, ni lo necesitó para controlar la epidemia, el conocimiento microbiológico (posterior) mejoró las medidas de prevención y tratamiento de este mal. Además, que se busquen los microhechos biológicos que subyacen a los macrohechos epidemiológicos es independiente, y compatible con que se busquen sus mecanismos sociales, las causas de las causas. No hay antagonismo entre epidemiología de factores de riesgo, epidemiología molecular y ecoepidemiología.

Por último, el criterio de plausibilidad está relacionado con el de consistencia o, mejor dicho, con el de repetibilidad. Al fin y al cabo, considerar los mecanismos equivale a examinar si la hipótesis causal se mantiene, y es consistente con circunstancias evidenciales biológicas, que son tests más severos y persuasivos de la hipótesis. Además, este criterio está relacionado con la generalización: considerar los mecanismos conocidos o postulados de una asociación ayuda a decidir si es aplicable a otros grupos de sujetos.

3.2.6. Especificidad y analogía

Son dos criterios despreciados por algunos autores. La especificidad postula que una causa conduce a un efecto único, no a efectos múltiples. Sin embargo, muchos factores aislados suelen tener muchos efectos (por ejemplo, el tabaco). En todo caso, cierta especificidad en la magnitud de la asociación puede contribuir a la verosimilitud de una hipótesis causal (así, en la relación tabaco/cáncer de pulmón el RR aumenta mucho más que en la del tabaco con otras enfermedades). La analogía postula que asociaciones causales similares podrían producir enfermedades similares. Por ejemplo, si un medicamento puede causar malformaciones congénitas, quizás puede hacerlo otro. Este criterio, contemplado por Hill y MacMahon, forma parte del concepto global de plausibilidad, pues se refiere a que la relación es análoga a alguna otra relación bien establecida; y, más bien, es útil para formular hipótesis más elaboradas que para evaluarlas.

3.2.7. El experimento

También llamado cesación de la exposición. Si un factor tiene un efecto, dicho efecto debería cesar cuando es removido el factor. El proyecto en marcha "Women's Health Initiative", que contrasta entre otros los efectos de una dieta baja en grasas sobre la incidencia del cáncer de mama, es un buen ejemplo. Este criterio no siempre es obtenible en poblaciones humanas, pero si el experimento se puede llevar a cabo constituye la prueba más sólida de causalidad. El criterio de experimentación es un criterio o regla metodológica de severidad o robustez de la hipótesis. Pero incluso puede ser visto, más que como un criterio, como un test directo de la hipótesis causal.

3.2.8. Coherencia

Para Susser y otros la coherencia incluye al criterio de analogía y plausibilidad. La coherencia dice que la asociación no debe entrar en conflicto con lo que se sabe de la enfermedad. Es decir, debe ser consistente con el conocimiento existente. Así, la asociación del cáncer de pulmón con el consumo de tabaco es coherente con el aumento temporal habido en las dos variables desde principios de siglo, y con la diferencia sexual en la mortalidad. Contribuyen a la coherencia las alteraciones histopatológicas observadas en el epitelio bronquial de los fumadores, y la presencia en el humo del tabaco de carcinógenos para la piel de los animales de laboratorio.

Pero la no coherencia (distinguirla del conflicto) no debería ser prueba en contra de la causalidad. Así, la ausencia de reproducción de la hipótesis en animales de laboratorio no va en contra de la relación en humanos. Y si el factor causa solo una pequeña proporción de la enfermedad total, la influencia abrumadora de otros factores puede hacer al patrón global incoherente.

Es de interés añadir que, volviendo a las consideraciones filosóficas de la causalidad, para nosotros los criterios de causalidad, como otras muchas reglas de la ciencia, no fueron concebidos como reglas de inferencia lógica, ni inductiva ni deductiva, sino como reglas metodológicas, reglas de discusión crítica, es decir criterios que ayudan, no imponen con fuerza lógica (o probable), a decidir si la evidencia apoya o rechaza, o no permite concluir, siempre tentativamente, sobre la hipótesis causal.

Precisamente que los criterios sean reglas metodológicas, decisionales, explica que puedan tener, como hemos dicho, excepciones. Las decisiones se incluyen en el método crítico de la ciencia, pero siempre como decisiones tentativas y sujetas a la crítica. Nuestras decisiones no tienen por qué ser decisiones últimas. Realmente, si se acepta que tienen esta naturaleza, sobraría tener que justificar que no son necesarios ni suficientes para la causalidad. Por eso ésta es una cuestión de juicio, gradual, aunque dicotómica en la acción (se actúa -o no-, como si la hipótesis causal fuese cierta). En todo caso, sería preferible evitar el término "criterios" por su connotación de necesarios y concluyentes (criterio = método definitivo o seguro de decisión). Se trata más bien de reglas o normas flexibles, orientadoras, y provisionales.

3.3. Otros criterios de causalidad

Aceptando que los criterios de Hill son útiles en la práctica, no capturan todos los elementos útiles para la inferencia causal. Por ello, recientemente diversos comités de expertos y autores a título individual han aumentado, modificado o desechado los criterios de Hill. Veamos algunos de esos otros criterios.

3.3.1. Descarte de explicaciones alternativas. Criterio de independencia

Puede haber explicaciones de la realidad alternativas a la causal, como son los sesgos y los FC y el azar. Algunos autores abogan por una discusión específica de estos factores (lo que en la práctica siempre se hace, por cierto) y no considerarlos solo indirectamente a través de algunos de los criterios de Hill.

Para muchos autores estos criterios "negativos" (el descarte razonable de sesgos y FC apoya la causalidad) deben ser la primera consideración en la evaluación de una hipótesis causal. El orden de la valoración debe ser: primero los sesgos y luego FC, pues si hay sesgo severo, ninguna manipulación de los datos salvará el problema. La valoración del azar vendrá en tercer lugar.

Una concreción de la consideración de los FC es el llamado por algunos autores criterio de "independencia". Este criterio es considerado en muchos estudios bajo la forma general de que la asociación no debe ser debida enteramente al efecto de un FC. El *U.S. Pooling Project* (1978) de estudios observacionales prospectivos sobre determinantes de la enfermedad coronaria lo incluyó entre sus criterios de causalidad: un factor de riesgo es independiente si mantiene al menos un cierto grado de efecto una vez controlada (generalmente por métodos multivariados) la influencia de otros determinantes relevantes de la enfermedad. Este criterio lo cumplían en 1978 el aumento de colesterol sérico, el consumo de tabaco y la hipertensión arterial (tras ajustar por edad y los otros factores predictores).

3.3.2. Poder explicativo

Este criterio tiene que ver con la audacia, originalidad, novedad, contenido informativo, rendimiento o poder predictivo o explicativo, interés e importancia de una hipótesis (causal). Son características interrelacionadas que sirven para evaluar *a priori* una hipótesis causal, ayudando así al progreso de la ciencia si esa hipótesis resulta corroborada *a posteriori*.

Una hipótesis es "audaz" si es nueva, va más allá de la vigente, predice hechos previamente insospechados; o si da respuesta a una pregunta no contestada todavía o que lo ha sido de forma contradictoria o no replicadamente. MacMahon resaltaba que en epidemiología, como en otras ciencias, el progreso en la búsqueda de asociaciones causales se favorece con hipótesis nuevas o más específicas. Popper distingue dos componentes de la audacia. Audacia de predecir de la hipótesis aspectos observables hasta ahora pasados por alto, que son tan osados que corren el riesgo de ser falsos y, por tanto ha de intentarse refutarlos. Por eso, si resisten los tests (criterios), se admite provisionalmente la hipótesis como verosímil: es muy improbable que el éxito de una teoría tan audaz, tan improbable, tan poderosa sea debido a errores o sesgos. La segunda audacia es suponer la realidad escondida tras lo observable: postular los mecanismos o las leyes biológicas subyacentes a la hipótesis o relación epidemiológica; se relaciona con la plausibilidad biológica.

La audacia, originalidad, novedad o interés de la hipótesis *a priori*, se relacionan con su poder para predecir o explicar más hechos observables novedosos e interesantes que la hipótesis previa, vigente, o que una rival. Un rico contenido informativo de la hipótesis permite que la contrastabilidad de la misma sea independiente y, por tanto, más severa, más fértil, más corroborable. Pero para que se produzca progreso en el conocimiento, algunas de esas nuevas predicciones deben ser corroboradas de hecho, superar contrastaciones ingeniosas y rigurosas. Esas pruebas rigurosas presuponen un alto grado de contrastabilidad o contenido *a priori*. De ahí la importancia del criterio llamado por Susser rendimiento predictivo: nuevas predicciones que resulten exitosas, concepto tomado de Popper.

Un ejemplo de teoría audaz es la de Ross sobre la aterosclerosis (1976), según la cual la placa de ateroma es la consecuencia de la respuesta local a la lesión endotelial. La lesión endotelial daba lugar a la agregación plaquetaria y el factor de crecimiento derivado de las plaquetas originaba una proliferación y emigración del músculo liso que producía la placa de ateroma. Incorporaba, pues, la plaqueta como nuevo elemento a la teoría clásica y subrayaba el fondo inflamatorio de la proliferación. Posteriormente (1986, 1993), nuevos hallazgos cambiaron la plaqueta por el monocito como célula fundamental. La teoría actual es una síntesis de las tres hipótesis iniciales: la inflamatoria de Virchow, la trombótica de Rokitansky y la lipídica de Anitschkow. El hecho clave de la aterosclerosis, particularmente de su comienzo, es el acúmulo de lipoproteínas; la trombosis es fundamental en la placa rota inestable y el infarto de miocardio, y la inflamación es evidente en el papel de monocitos, macrófagos y proliferación muscular lisa en la génesis del ateroma. La moderna teoría subsume, pues, a las anteriores, predice los hechos previstos por ellas y, además, añade nuevas predicciones; es más rica en contenido (y verosímil tras contrastarse).

El interés de una teoría depende tanto de la sorpresa como de la importancia. El interés teórico es el potencial para cambiar lo que los científicos creen sobre importantes relaciones causales. Lo sorprendente, lo contra-intuitivo debería ser un criterio importante para las buenas hipótesis. El interés decae ante las numerosas replicaciones de una hipótesis; el interés se reaviva al buscar y encontrar factores contextuales que modifiquen el efecto. La importancia se relaciona con la frecuencia, gravedad y transcendencia social del problema implicado; los beneficios y aplicabilidad anticipables de su resolución; o el avance del conocimiento en ese u otros campos

y las futuras guías de investigación que probablemente se deriven. Por ejemplo, las hipótesis sobre el cáncer son más importantes que aquellas sobre varices porque son más las personas (y los fenómenos biológicos y psicológicos) que están afectadas más profundamente por el primero. Si bien, en ejemplos menos obvios las ramificaciones de una teoría suelen ser difíciles de anticipar.

Es fácil deducir la conexión de este criterio con los de consistencia, plausibilidad, independencia y con la generalización.

3.3.3. Calidad de la evidencia

Este criterio se refiere a la categorización de la evidencia por la calidad de sus fuentes o tipos de estudios. Se relaciona con la validez interna y la fuerza "demostrativa" de causalidad derivada de los tipos de estudios o fuentes suministradoras de la evidencia.

Aunque considerado en parte en el criterio de consistencia (replicación en diferentes métodos o diseños), la acepción moderna del criterio de calidad de la evidencia introduce la clasificación de los tipos o diseños en orden descendente según su fortaleza, es decir, del grado en que éstos consiguen el control de variables extrañas. Así, diversos grupos han establecido clasificaciones jerárquicas de la calidad de la evidencia científica, fundamentalmente de cara a priorizar intervenciones preventivas o a evaluar tecnologías sanitarias (tabla 13-3).

Tabla 13-3. Clasificación de la calidad de la evidencia según el diseño del estudio fuente.

Diseño
Ensayo clínico
Estudio de cohortes
Estudio de casos y controles
Estudio transversal
Estudio de riesgo agregado
Serie de casos
Informe de un caso

Los tipos de estudio están enumerados en orden descendente de calidad. Estos o similares criterios son utilizados por los grupos de trabajo americano y canadiense de servicios clínicos preventivos, y diversas agencias de evaluación de tecnologías sanitarias, para juzgar la calidad de evidencia en la que apoyar algunas recomendaciones de práctica clínica.

La corroboración de una hipótesis causal es mayor cuanto más severas son las condiciones de contrastación; en este caso, la homogeneización o minimización de factores perturbadores (típicamente FC) de la relación "causal" en el estudio contrastador, conseguible en mayor grado por la randomización en estudios experimentales; en un grado menor por restricción, equiparamiento, análisis estratificado o modelos matemáticos en estudios observacionales de cohortes o casos-control; a continuación, los estudios correlacionales poblacionales; y en un grado mínimo en estudios descriptivos, series de casos o experiencias clínicas.

3.4. Un ejemplo de aplicación de criterios de causalidad

Veamos cómo se puede practicar la inferencia causal sobre la relación entre la infección por *Helicobacter pylori* (*H. pylori)* y la úlcera duodenal utilizando algunos de los principales criterios (básicamente los de Hill).

1. Relación temporal. En un estudio de 454 pacientes sometidos a endoscopia, 34 de los 321 pacientes positivos al *H. pylori* (11%) desarrollaron posteriormente úlcera duodenal a lo largo de 10 años, comparados con 1 de 133 *H. pylori*-negativos (0,8%).
2. Fuerza de la relación. El anterior estudio produce un RR próximo a 14.

3. Relación dosis-respuesta. El riesgo de úlcera duodenal aumenta con la densidad de *H. pylori* por mm^2 de mucosa gástrica.
4. Replicación de los hallazgos. Muchas de las observaciones referidas al *H. pylori* han sido replicadas repetidamente.
5. Plausibilidad biológica. El *H. pylori* tiene sitios de unión sobre las células antrales y puede seguir a estas células en el duodeno. El *H. pylori* también induce mediadores de inflamación. La mucosa infectada por *H. pylori* está debilitada y es susceptible a los efectos dañinos del ácido gástrico.
6. Consideración de explicaciones alternativas. La asociación del *H. pylori* con la úlcera duodenal podría explicarse por su fuerte asociación con el consumo de tabaco. Sin embargo, los datos sugieren que el tabaco puede incrementar el riesgo de úlcera duodenal en pacientes infectados por *H. pylori* pero no es un factor de riesgo en pacientes en quienes el *H. pylori* ha sido erradicado. Por ello, este gérmen es necesario para la producción de la úlcera por el tabaco, y ejerce al menos un papel co-causal.
7. Cesación de la exposición. La erradicación del *H. pylori* cura la úlcera duodenal en la misma medida que los antagonistas de los receptores de histamina. Las tasas de recurrencia de úlcera a largo plazo tras erradicar al *H. pylori* con terapia triple antimicrobiana fueron cero, comparadas con tasas de recaida del 60-80% encontradas en pacientes tratados con antagonistas de los receptores de histamina.
8. Especificidad de la asociación. La prevalencia de *H. pylori* en pacientes con úlcera duodenal es 90-100%. Sin embargo, el *H. pylori* se ha encontrado en algunos pacientes con úlcera gástrica e incluso en individuos asintomáticos.
9. Consistencia con otro conocimiento. La prevalencia de infección por *H. pylori* es la misma en varones que en mujeres. La incidencia de úlcera duodenal, que años atrás se creía que era mayor en varones, es igual en años recientes. Además, se cree que la prevalencia de la enfermedad ulcerosa alcanzó su pico en la última parte del siglo XIX, y la prevalencia de *H. pylori* puede haber sido mucho mayor en esa época debido a las pobres condiciones de vida. Este razonamiento está también basado en observaciones actuales de que la prevalencia de *H. pylori* es mucho mayor en los países en vías de desarrollo.

Aunque los anteriores criterios no permiten cuantificar si la asociación es o no causal, permiten sugerir de forma fuerte que lo es. En otras palabras, la hipótesis del *H. pylori* como agente causal de la úlcera duodenal parece razonablemente corroborada, hoy por hoy.

Sin embargo, los criterios de causalidad no son rígidas exigencias que permiten decidir si la asociación es o no causal. Entre otras razones porque estos criterios no son completos: hemos de considerar también, como ya dijimos, los posibles sesgos y la calidad de las fuentes generadoras de las evidencias. Además, tienen excepciones y no hay acuerdo sobre cuáles deben primar o como operativizarlos o valorarlos (¿es lo mismo un RR=2 en epidemiología nutricional que en la ocupacional?), o en qué medida se cumplen. Pero a pesar de sus limitaciones, los criterios, como directrices orientadoras, facilitan la evaluación de la evidencia y eventualmente ayudan a tomar una decisión (tratar o no la infección por *H. Pylori* en la úlcera duodenal, si se dispone, como ocurre, de tratamiento eficaz).

4. LA PRÁCTICA DE LA INFERENCIA CAUSAL

Al examinar la práctica de la inferencia causal en epidemiología, Weed encuentra un uso relativamente pobre y heterogéneo de los criterios de causalidad de Hill, por lo que propugna

un mayor uso de estos criterios, estudiar sus relaciones y desarrollar otros. Dado un determinado conjunto de criterios, no está claro como combinarlos o priorizarlos, ni si esto es realmente necesario o factible. En una aproximación muy intuitiva, y a raíz de un reciente debate sobre los límites de la epidemiología como ciencia no experimental, se llegó a comentar que no son creíbles, de entrada, asociaciones cuyos RR no sean superiores a 3 o 4. Otros autores (Trichopoulos y MacMahon) apuestan por hacer jugar un papel esencial a la plausibilidad biomédica ante efectos inconsistentes y generalmente débiles, y apostar por la causalidad ante asociaciones fuertes y consistentes, aun en ausencia de apoyo biológico. Estos mismos autores apuntan que ante resultados inconcluyentes, moderadamente compatibles con la hipótesis, puede ajustarse la principal medida de efecto no solo por confusores claros sino también por variables con cuestionable, incluso improbable, influencia confusora. Esto mejora el control de la confusión y la consiguiente ampliación del intervalo de confianza reduce la tentación de sobreinterpretación.

En la práctica de la epidemiología y salud pública, se empieza también a considerar conjuntamente los criterios de Hill y los de calidad de la evidencia, es decir, a ponderar los criterios de evidencia a la luz de los diseños de las fuentes de la evidencia, considerando adicional y específicamente los sesgos y FC. Algunos autores sitúan en un primer escalón de evidencia a los estudios de mayor peso de diseño (ensayos clínicos y estudios de cohorte), en segundo nivel a RR grandes y consistentes, y a los estudios casos-control y series temporales, y en el último los RR pequeños aunque plausibles y a los estudios transversales.

Sin embargo, prácticamente siempre el juicio causal se realiza valorando evidencias de muchos tipos de estudios. Así, los efectos del consumo de cigarrillos han sido demostrados mediante experimentos con animales, grupos controlados y encuestas con seres humanos. De hecho, la investigación en humanos muchas veces comienza con pistas etiológicas de observaciones clínicas o estudios descriptivos, prosigue con estudios observacionales analíticos y, si éstos parecen corroborar los hallazgos, se realiza una evaluación experimental, cuando es posible (agentes potencialmente beneficiosos). Esta es una secuencia inversa a la evaluación de calidad de las evidencias de las escalas jerárquicas. La consideración conjunta de los tres grandes tipos de estudios disponibles, la variedad de diseños (no la jerarquización y prima excesiva de los experimentos), es, como vimos en el criterio de consistencia, una regla que realza la inferencia.

Como señala Weed, un problema adicional es que la selección, clasificación, definición e interpretación de criterios varían de usuario a usuario y de grupo a grupo, lo que puede producir juicios muy diferentes sobre la causación a partir de la misma evidencia. De hecho, se suele decir que la valoración causal es materia de juicio subjetivo. De juicio sí; subjetivo, no tanto. A pesar de la aparentemente subjetiva disparidad, el proceso inferencial se objetiva al abrirse la discusión a los demás. Las razones en pro y en contra se pueden sopesar y discutir; así se posibilita el acuerdo. El refinamiento de los criterios y su sistematización, su mayor reproducibilidad por todos, el mayor consenso sobre su definición y reglas de interpretación, mejorará la toma de decisiones causales, pero ésta es siempre provisional, revisable: todos podemos errar en la misma dirección, o la evidencia puede cambiar.

Dada la cantidad de criterios existentes y la diversidad de su manejo, en la tabla 13-4 se listan los criterios más importantes, en un orden razonable según diversas perspectivas. Se prioriza el descarte razonable de explicaciones o factores no causales. El cuadro distingue también los criterios que afectan a la validez interna, a la validez externa y a la comparación con otra evidencia. En general, los criterios más fuertes son la precedencia temporal, la magnitud de la asociación, la consistencia con otros hallazgos y la credibilidad biológica de la hipótesis.

Tabla 13-4. Cuestiones más relevantes para valorar la causalidad

A. Validez interna - Explicaciones no causales (o incluso ausencia de asociación)
1. Sesgos
2. Factores de confusión
3. Azar
B. Validez interna - Características positivas
4. Relación temporal
5. Fuerza
6. Dosis-respuesta
7. Consistencia
C. Validez externa - Generalizabilidad
8. Aplicación a otras poblaciones
D. Comparación con otra evidencia
9. Calidad de la evidencia
10. Consistencia
11. Plausibilidad

Modificado de Elwood (1988).

Por último, las cuestiones teóricas son independientes de si el epidemiólogo, como científico, debe aconsejar o no la acción, basándose en las evidencias científicas de causalidad. Su decisión puede quedar "limitada" a su evaluación (metodológica) de la hipótesis causal, basada en la evidencia científica. Si no decide él sobre la acción, no faltarán otros, científicos o políticos, que, con evidencias científicas u otras consideraciones, tomarán la decisión práctica en nombre de todos.

5. REFERENCIAS BIBLIOGRÁFICAS

ABELSON RP. *Statistics as principal argument.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

BANEGAS JR, RODRÍGUEZ ARTALEJO F, REY CALERO J. *Popper y el problema de la inducción en epidemiología.* Rev Esp Salud Púb 2000;74:327-339.

BUNGE M. *La investigación científica. Su estrategia y su filosofía.* Barcelona: Ariel, 1985.

ELWOOD JM. *Causal relationships in Medicine.* Nueva York: Oxford University Press, 1988.

GORDIS L. *Epidemiology.* Philadelphia: W.B. Saunders, 1996.

GREENLAND S (editor). *Evolution of epidemiologic ideas.* Chesnut Hill, MA: E.R.I., 1987.

HILL AB. *The environment and disease: association or causation?* Proc Roy Soc Med 1965; 58: 295-300.

MACMAHON B, TRICHOPOULOS D. *Epidemiology. Principles and methods.* 2ª ed. Boston: Little, Brown, 1996.

POPPER KR. *The logic of scientific discovery.* 3ª ed. London: Routledge, 1980.

RODRÍGUEZ ARTALEJO F, BANEGAS JR, RODRÍGUEZ ARTALEJO C, RODRÍGUEZ ARTALEJO A. *Principios del diseño experimental: del laboratorio a la epidemiología.* Revisiones en Salud Pública 1995; 4: 149-168.

RODRÍGUEZ-ARTALEJO F, BANEGAS JR, HERRUZO R, REY-CALERO J. *Causalidad en Epidemiología.* En: Piédrola Gil. Medicina Preventiva y Salud Pública. Barcelona: Masson, 2001.

ROTHMAN KJ (Ed.). *Causal Inference.* Chesnut Hill, MA: E.R.I., 1988.

ROTHMAN KJ, GREENLAND S. *Causation and causal inference.* En: Rothman KJ, Greenland S. Modern Epidemiology. 2ª ed. Philadelphia: Lippincott-Raven, 1998.

STUART MILL J. *De los cuatro métodos de indagación experimental.* Cuadernos Teorema. Valencia: G. Lormo, 1980.

SUSSER M. *What is a cause and how do we know one*? *A grammar for pragmatic epidemiology.* Am J Epidemiology 1991; 133: 635-648.