Mauricio Hernández Ávila

# Epidemiología

## *Diseño y análisis de estudios*



Instituto Nacional de Salud Pública

Editorial Médica Panamericana

# Epidemiología

## *Diseño y análisis de estudios*

# Epidemiología

## *Diseño y análisis de estudios*

Editor
Mauricio Hernández Ávila

Médico Cirujano por la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM). Diplomado en Estadística por el Instituto de Investigación en Matemáticas Aplicadas y Sistemas de la UNAM. Maestría y Doctorado en Epidemiología por la Escuela de Salud Pública de la Universidad de Harvard. Profesor Investigador en Ciencias Médicas F-Visitante del Instituto Nacional de Salud Pública. Miembro de comités consultivos de investigación de México y del Institute of Medicine de Estados Unidos. Miembro de la Academia Nacional de Medicina y de la Academia Mexicana de Ciencias. Investigador Nacional nivel III del Sistema Nacional de Investigadores.



Instituto Nacional
de Salud Pública


BUENOS AIRES • BOGOTÁ • CARACAS • MADRID • MÉXICO • SAO PAULO

**www.medicapanamericana.com**

Título:
Epidemiología. Diseño y análisis de estudios.
Mauricio Hernández Ávila


EDITORIAL MEDICA panamericana

**Argentina**
Editorial Médica Panamericana S.A.

Marcelo T. de Alvear 2145 (1122) Buenos Aires Argentina
Tel.: (54-11)4821-2066 / 5520/ Fax: (54-11) 4821-1214
info@medicapanamericana.com

**España**
Editorial Médica Panamericana S.A.

Alberto Alcocer24 (28036) Madrid, EspañaTel.: (34) 91-1317800 / Fax: (34) 91-1317805 / (34) 91-4570919
info@medicapanamericana.es

**Colombia**
Editorial Médica Internacional, LTDA

Carrera7a A No.69-19 Santa Fe de Bogotá DC.
Tel.: (57-1)345-4508 / 34-5014/ Fax:(57-1) 314-5015/ 345-0019
infomp@medicapanamericana.com.co

**México**
Editorial Médica Panamericana S.A.de C.V.

Hegel No 141 2do. piso, Chapultepec Morales
C.P. 11570 México D.F.
Tel.: (5255) 5250-0664 / 5262 9470 / Fax: (5255) 2624-2827
infomp@medicapanamericana.com.mx

**Venezuela**
Editorial Médica Panamericana, C.A.

Edificio Polar, Torre Oeste, Piso7 Of. 7-A
Plaza Venezuela. UrbanizaciónLos Caobos
Parroquia El Recreo, Municipio Liberador Caracas D.F.
Tel.: (58-212) 793-2857/ 6906 / 5985 / 1666
Fax: (58-212) 793-5885
info@medicapanamericana.com.ve

Visite nuestra página Web:
www.medicapanamericana.com

ISBN: 978-968-7988-67-2


ISBN: 978-970-9874-04-4


Impreso en México / Printed in Mexico

Esta edición se terminó de imprimir en el mes de _____ en los talleres de _____.
Se tiraron _____ ejemplares más sobrantes para su reposición.

# Coeditores

### Epidemiología

**Eduardo Lazcano Ponce**
Médico Cirujano por la Universidad Autónoma de Puebla y Especialista en Medicina Familiar por el Instituto Politécnico Nacional. Maestro en Ciencias en Epidemiología por la Universidad Nacional Autónoma de México. Doctor en Ciencias en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Posdoctorado en la International Agency for Research on Cancer, en Lyon, Francia. Investigador en Ciencias Médicas F del INSP. Miembro del Sistema Nacional de Investigadores, de la Academia Nacional de Medicina y de la Academia Mexicana de Ciencias.

**Sergio López Moreno**
Médico Cirujano y Partero por el Instituto Politécnico Nacional. Especialista en Medicina Familiar por la Secretaría de Salud y Asistencia. Doctor en Ciencias en Epidemiología por la Universidad Nacional Autónoma de México. Profesor-investigador en el Departamento de Atención a la Salud de la Universidad Autónoma Metropolitana-Xochimilco. Miembro del Sistema Nacional de Investigadores.

### Análisis estadístico

**Martha María Téllez Rojo Solís**
Matemática con especialidad y maestría en Estadística por la Universidad Nacional Autónoma de México. Doctora en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Profesor titular de Bioestadística en los programas de maestría y doctorado y coordinadora del programa de Doctorado en Ciencias con área de concentración de Epidemiología del INSP. Miembro del Sistema Nacional de Investigadores.

### Selección y diseño de ejercicios

**Gabriela Torres Mejía**
Médica Cirujana por la Universidad La Salle. Maestra en Ciencias de la Salud con área de concentración en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Doctorado en Epidemiología por la London School of Hygiene & Tropical Medicine. Investigador en Ciencias Médicas C del INSP.

### Coordinación editorial

**Carlos Oropeza Abúndez**
Licenciado en Ciencias Políticas y Administración Pública por la Universidad Nacional Autónoma de México. Subdirector de Publicaciones y Comunicación Científica del Instituto Nacional de Salud Pública. Editor Ejecutivo de la revista *Salud Pública de México.*

### Revisión técnica

En la revisión técnica de los capítulos que corresponden al diseño de estudios epidemiológicos participaron Martha Ethelia López, Javier Idrovo Velandia, Adolfo Hernández y Alexánder Corcho. La revisión técnica de los capítulos de análisis estadístico estuvo a cargo de Héctor Lamadrid y Aarón Salinas.

# Colaboradores

**Angélica Rocío Ángeles Llerenas**
Médica Cirujana y Partera por la Benemérita Universidad Autónoma de Puebla (BUAP). Diplomada en Investigación Clínica por la BUAP. Maestra en Ciencias de la Salud, área de concentración en Epidemiología por el Instituto Nacional de Salud Pública de México (INSP). Diplomada en Bioética por la Facultad Latinoamericana de Ciencias Sociales (FLACSO), Buenos Aires, Argentina. Profesora adjunta del Seminario de Epidemiología I y Profesor titular del Seminario de Investigación en Epidemiología, INSP.

**Juan José Calva Mercado**
Subdirector de Investigación Clínica en el Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán.

**Esteve Fernández Muñoz**
Licenciado en Medicina y Cirugía por la Universidad Autónoma de Barcelona, con formación en Epidemiología y Salud Pública en la misma universidad y en el Instituto Municipal de Investigación Médica (Máster en Salud Pública y Doctor en Medicina). Investigador del Instituto Catalán de Oncología (Barcelona, España) y director de la Unidad de Investigación en Tabaquismo. Profesor Asociado de la Universidad Pompeu Fabra (Barcelona, España) en el Programa de Maestría y Doctorado en Salud Pública. Editor de Gaceta Sanitaria, la revista científica de la Sociedad Española de Salud Pública y Administración Sanitaria.

**Ma. de Lourdes Guadalupe Flores Luna**
Química Farmacobióloga por la Universidad Autónoma de San Luís Potosí. Maestra y Doctora en Ciencias de la Salud en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Coordinadora de la Maestría en Ciencias de la Salud en Bioestadística del INSP. Miembro del Sistema Nacional de Investigadores.

**Francisco Garrido Latorre**
Médico Cirujano por el Instituto Superior de Ciencias Médicas de la Habana, Cuba. Maestro en Ciencias en Epidemiología y Doctor en Salud Pública por el Instituto Nacional de Salud Pública de México. Director General de Evaluación del Desempeño de la Secretaría de Salud.

**Pedro Gutiérrez Castrellón**
Médico Cirujano por la Facultad de Medicina de la Universidad Juárez del Estado de Durango. Especialista en Pediatría por el Instituto Nacional de Pediatria (INP) y la Universidad Nacional Autónoma de México (UNAM), Subespecialista en Urgencias Pediátricas (INP, UNAM). Maestro y Doctor en Ciencias Médicas, UNAM. Director de Investigación (INP). Miembro del Sistema Nacional de Investigadores. Coordinador del Centro de Análisis de la Evidencia en Pediatría INP-COCHRANE.

### Mauricio Hernández Ávila

Médico Cirujano por la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM). Diplomado en Estadística por el Instituto de Investigación en Matemáticas Aplicadas y Sistemas de la UNAM. Maestría y Doctorado en Epidemiología por la Escuela de Salud Pública de la Universidad de Harvard. Profesor Investigador en Ciencias Médicas F-Visitante del Instituto Nacional de Salud Pública (INSP). Miembro de comités consultivos de investigación de México y del Institute of Medicine de Estados Unidos. Miembro de la Academia Nacional de Medicina y de la Academia Mexicana de Ciencias. Investigador Nacional nivel III del Sistema Nacional de Investigadores. Subsecretario de Prevención y Promoción de la Salud de la Secretaría de Salud.

### Adolfo Gabriel Hernández Garduño

Médico Cirujano por la Universidad Nacional Autónoma de México (UNAM). Especialista en Pediatría por el Hospital General de México. Maestro en Ciencias de la Salud con área de concentración en Nutrición y Doctor en Ciencias en Salud Pública por el Instituto Nacional de Salud Pública (INSP). Profesor titular en la Maestría en Ciencias en Epidemiología Clínica (INSP). Miembro del Sistema Nacional de Investigadores.

### Bernardo Hernández Prado

Licenciado en Psicología Social por la Universidad Autónoma Metropolitana-Iztapalapa. Maestro en Psicología Social por la London School of Economics and Political Sciences, Universidad de Londres. Doctor en Salud y Comportamiento Social por la Harvard School of Public Health, Universidad de Harvard. Investigador en Ciencias Médicas E del Instituto Nacional de Salud Pública. Investigador nacional nivel I.

### Álvaro Javier Idrovo Velandia

Médico Cirujano y Magíster en Salud Pública de la Universidad Nacional de Colombia. Especialista en Higiene y Salud Ocupacional de la Universidad Distrital en Bogotá, Colombia. Maestro en Ciencias en Salud Ambiental y Doctor en Ciencias en Epidemiología en el Instituto Nacional de Salud Pública. Profesor de la Facultad de Enfermería y Nutriología de la Universidad Autónoma de Chihuahua. Miembro del Sistema Nacional de Investigadores.

### Pablo Antonio Kuri Morales

Médico Cirujano y Maestro en Ciencias Sociomédicas con énfasis en Epidemiología por la Universidad Nacional Autónoma de México. Diplomado en Gerencia en Salud Pública en la Universidad de Emory, EUA. Certificado por el Consejo Nacional de Salud Pública. Investigador titular A del Sistema Nacional de Investigadores y miembro de la Academia Mexicana de Cirugía y de la Sociedad Mexicana de Salud Pública. Director General del Centro Nacional de Vigilancia Epidemiológica y Control de Enfermedades de la Secretaría de Salud.

### Héctor Manuel Lamadrid Figueroa

Médico Cirujano por la Universidad Autónoma del Estado de Morelos. Maestro en Ciencias con área de concentración en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Candidato a Doctor en Ciencias de la Salud con concentración en Epidemiología por el INSP. Investigador en Ciencias Médicas B en el INSP. Profesor titular de las materias Bioestadística Básica (2004-2005) y Métodos Intermedios de Bioestadística (2004-2005) en el INSP.

**Eduardo Lazcano Ponce**
Médico Cirujano por la Universidad Autónoma de Puebla, especialista en Medicina Familiar por el Instituto Politécnico Nacional. Maestro en Ciencias en Epidemiología por la UNAM. Doctor en Ciencias en Epidemiología por el Instituto Nacional de Salud Pública. Posdoctorado en el International Agency for Research on cancer, en Lyon, Francia. Investigador titular del INSP. Miembro del Sistema Nacional de Investigadores, de la Academia Nacional de Medicina y de la Academia Mexicana de Ciencias. Director del Centro de Investigación en Salud Poblacional.

**Sergio López Moreno**
Médico Cirujano y Partero por el Instituto Politécnico Nacional. Especialista en Medicina Familiar y Comunitaria por la Universidad Nacional Autónoma de México. Profesor titular en los programas de Maestría en Medicina Social y Doctorado en Salud Colectiva de la División de Ciencias Biológicas y de la Salud de la Universidad Autónoma Metropolitana-Xochimilco. Editor de Sociomedicina de la revista *Salud Pública de México*. Miembro del Sistema Nacional de Investigadores.

**Fernando Meneses González**
Médico Cirujano y Partero por la Escuela Superior de Medicina del Instituto Politécnico Nacional. Especialista en Epidemiología Aplicada (Secretaría de Salud/Centro de Control de Enfermedades, EUA) y Maestro en Ciencias de la Salud en el Trabajo (Universidad Autónoma Metropolitana). Alumno del programa de Doctorado en Epidemiología de la Facultad de Medicina en la Universidad Nacional Autónoma de México. Miembro del Sistema Nacional de Investigadores. Miembro de la Academia Nacional de Medicina. Miembro de la Academia Mexicana de Ciencias.

**Alejandra Moreno Altamirano**
Cirujana Dentista y Maestra en Ciencias en Epidemiología por la Universidad Nacional Autónoma de México (UNAM). Profesora titular del seminario de Epidemiología, programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud en la UNAM. Profesora de Carrera de tiempo completo en la Facultad de Medicina de la UNAM.

**Esteban Rodríguez Solís**
Secretaría de Salud

**Marcos Adán Ruiz Rodríguez**
Secretaría de Salud

**Eduardo Salazar Martínez**
Médico Cirujano por la Universidad Nacional Autónoma de México, con especialidad en Medicina Familiar por el Instituto Politécnico Nacional y el Instituto Mexicano del Seguro Social. Maestro en Ciencias de la Salud con área de concentración en Epidemiología y Doctor en Ciencias de la Salud con énfasis en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Posdoctorado en el Departamento de Nutrición de la Escuela de Salud Pública de la Universidad de Harvard, como investigador visitante. Investigador Nacional nivel I. Investigador en Ciencias Médicas E del INSP.

**Daniela Sotres Álvarez**
Licenciatura en Matemáticas Aplicadas por el Instituto Tecnológico Autónomo de México. Maestra en Ciencias en Bioestadística por la Universidad de Carolina del Norte en Chapel Hill. Jefe del Departamento de Análisis y Procesamiento de Datos del Centro de Investigación en Nutrición y Salud del Instituto Nacional de Salud Pública (INSP) del 2002 al 2005. Investigador en Ciencias Médicas B. Profesor titular de Métodos Intermedios de Bioestadística en el INSP.

**Martha María Téllez Rojo Solís**
Matemática con especialidad y Maestra en Estadística por la Universidad Nacional Autónoma de México. Doctora en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Profesora titular de Bioestadística en los programas de maestría y doctorado y coordinadora del programa de Doctorado en Ciencias con área de concentración de Epidemiología del INSP. Miembro del Sistema Nacional de Investigadores. Directora de Ecología Humana del Centro de Investigación en Salud Poblacional del INSP.

**Gabriela Torres Mejía**
Médica Cirujana por la Universidad La Salle. Maestra en Ciencias de la Salud con área de concentración en Epidemiología por el Instituto Nacional de Salud Pública (INSP). Doctorado en Epidemiología por la London School of Hygiene & Tropical Medicine. Investigador en Ciencias Médicas C del INSP. Jefa del Departamento de Análisis de Riesgos Emergentes del Centro de Investigación en Salud Poblacional del INSP.

**Eduardo Velasco Mondragón**
Médico Cirujano por la Facultad de Medicina de la UNAM. Especialidad en Medicina Familiar en el Instituto Politécnico Nacional. Maestro en Ciencias en Sistemas de Salud por la Escuela de Salud Pública de México. Doctor en Epidemiología con área de concentración en Enfermedades Infecciosas por la Escuela de Salud Pública de la Universidad Johns Hopkins.

**José Luis Viramontes Madrid**
Médico Cirujano por la Universidad Nacional Autónoma de México. Especialista en Neumología (residencia en el Hospital General de México) y Maestro en Ciencias Médicas por la misma universidad. M.Sc. en Epidemiología Clínica y Economía de la Salud por la Universidad de McMaster en Hamilton, Canadá. Profesor titular de la maestría en Epidemiología Clínica del Instituto Nacional de Salud Pública y Director Asociado de Investigación Clínica en Merck Sharp & Dohme de México.

**José Salvador Zamora Muñoz**
Actuario por la Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM). Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas. UNAM.

# Índice

# Prólogo

La epidemiología ha sido considerada desde hace mucho tiempo, junto a la bioestadística, como una ciencia fundamental de la salud pública. Sus métodos se usan para monitorear la salud de las poblaciones e identificar problemas emergentes, para probar hipótesis concernientes a las causas de la enfermedad y para evaluar enfoques preventivos a través de experimentos cuidadosamente diseñados y de observaciones sobre los resultados de intervenciones en la población. Los métodos epidemiológicos tienen un margen de aplicación mayor en la investigación biomédica y son base de los métodos de investigación clínica e investigación de resultados (la aplicación de métodos cuantitativos a la investigación mediante el uso de grupos de sujetos con características específicas). Los métodos de la epidemiología proveen un puente que une a la investigación de laboratorio y pacientes con la población en general. Un ejemplo de esto son las implicaciones de los notables avances en la genética de la enfermedad, que requieren de enfoques epidemiológicos para utilizarlos en un contexto poblacional. En esta era en que las infecciones emergentes dan lugar a epidemias globales, se utilizan también los enfoques epidemiológicos para vigilar su aparición, encontrar sus causas y dirigir el desarrollo de intervenciones.

El punto de partida para la aplicación de la metodología epidemiológica en el ámbito de la salud pública y la investigación biomédica es una preparación sólida en los métodos que le son propios. *Epidemiología. Diseño y análisis de estudios* delinea estos métodos para los estudiantes y profesionales de la salud que buscan adquirir los fundamentos de la disciplina. La obra cubre, en primera instancia, la historia de la epidemiología, una ciencia que ha evolucionado constantemente para hacer frente tanto a los retos planteados por el cambio en los patrones de enfermedad —de enfermedades infecciosas a enfermedades crónicas— como a la incorporación del conocimiento genético y de marcadores complejos de riesgo y enfermedad temprana. Un capítulo posterior describe los indicadores fundamentales de la salud poblacional, los cuales necesitan ser comprendidos y utilizados por todos los actores involucrados en el monitoreo y mejora de la misma. Gran parte del libro cubre los fundamentos del diseño de estudios, conocimientos que son relevantes para todas las aplicaciones de la epidemiología y que reciben una cobertura adecuada y profunda. Una característica notable de la epidemiología es que sus principales métodos y diseños de estudio han cambiado poco durante los últimos 50 años, ya que el desarrollo posterior consistió en gran parte en la aplicación de sus enfoques estándares para el estudio de las enfermedades crónicas. Los últimos capítulos del libro cubren los métodos bioestadísticos de uso más frecuente para datos epidemiológicos, y complementan la cobertura que se hace de estos métodos en los cursos de bioestadística, colocando los modelos en un contexto epidemiológico.

Para aquellos que quieren interpretar y utilizar evidencia epidemiológica, estas bases pueden ser suficientes. Para aquellos otros que quieren convertirse en investigadores epidemiológicos independientes, *Epidemiología. Diseño y análisis de estudios* será un punto de partida que deberá ser complementado con cursos más avanzados en la materia, así como en bioestadística, y con la experiencia necesaria en la realización de investigaciones concretas. La epidemiología tiene actualmente muchas ramas distintas –genética, ambiental, de enfermedades infecciosas, del cáncer, cardiovascular, entre otras–, por lo que el estudiante que quiera especializarse en un tema de investigación necesitará extender aún más su campo de conocimiento y competencias.

Se requiere de autores experimentados para escribir libros de texto excelentes, como se muestra en los capítulos de esta obra de nivel intermedio. Aquí la lista incluye educadores, académicos e investigadores de varias instituciones mexicanas involucradas con la salud pública, quienes aportan al libro su propio trabajo en México, y con ello lo hacen particularmente relevante para los cursos que se imparten en este país. Pero los ejemplos del libro tienen un alcance incluso más amplio. En mi propia enseñanza, todavía me apoyo en estudios pioneros sobre el tabaquismo y la salud llevados a cabo en 1950 por Wynder y Graham y por Doll y Hill. Aunque terminados hace tiempo, las lecciones aprendidas en ellos son aplicables en muchas situaciones, al igual que las que se incorporan en este libro.

Las obras introductorias con frecuencia se vuelven hitos que dejan huella en los estudiantes de una disciplina. En la facultad de medicina donde estudié, usé la primera edición de *Epidemiology: principles and methods*, de MacMahon y Pugh, y como estudiante de epidemiología en la Escuela de Salud Pública de Harvard recurrí a la segunda edición. Todavía recuerdo la claridad del texto y sus explicaciones de los conceptos. En la Escuela Bloomberg de Salud Pública de Johns Hopkins hemos usado últimamente *Epidemiology*, de Leon Gordis (mi antecesor como titular del departamento), en nuestro curso introductorio: "Epidemiología 1", aunque antes utilizábamos *Fundamentals of epidemiology*, escrito originalmente por Abe Lilienfeld, quien precedió a Gordis en la cátedra. Los estudiantes me comentan frecuentemente sobre el texto de Gordis: "Su estilo es cálido y accesible, amistoso para el lector".

Los autores de *Epidemiología. Diseño y análisis de estudios* recibirán en el futuro el mismo tipo de comentarios y generarán la misma clase de recuerdos a medida que el libro se convierta en un estándar en países de habla hispana. Ofrezco mis felicitaciones a los autores, entre los cuales se encuentran amigos y colaboradores de hace mucho tiempo, por haber completado esta obra. Tengan por seguro que estaré atento para cuando aparezca la traducción al inglés.

Jonathan M. Samet
Director del Departamento
de Epidemiología
Johns Hopkins University

# Introducción

La epidemiología ha tenido un desarrollo conceptual y metodológico vertiginoso durante los últimos años. Actualmente no sólo se dedica al estudio de la distribución, frecuencia, determinantes, predicciones y control de los factores relacionados con la salud y la enfermedad de las poblaciones humanas sino, de manera quizás más importante, a la aplicación de sus resultados en beneficio de ellas.

La epidemiología es considerada la ciencia básica de la salud pública y la aplicación rigurosa de sus métodos constituye una fuente de información para la formulación de políticas de salud en el ámbito poblacional. La epidemiología estudia, sobre todo, las potenciales relaciones causales que se establecen entre la presencia de exposiciones determinadas y el desarrollo de enfermedades específicas, así como sus múltiples posibilidades de prevención. La epidemiología estudia la salud de los grupos humanos en relación con su medio ambiente.

Diversos ejemplos de la utilidad de la investigación epidemiológica han sido descritos desde hace más de 100 años. Inicialmente los métodos epidemiológicos surgieron del estudio de las epidemias de las enfermedades infecciosas, y posteriormente se convirtieron en la herramienta fundamental para generar estrategias de prevención y control una vez que se ha evaluado la asociación de diversas fuentes de exposición y enfermedad. Entre los muchos ejemplos que pueden describirse destacan el estudio de tabaquismo asociado a cáncer de pulmón, que ha dado lugar a la virtual prohibición de exposición a humo de tabaco ambiental en muchos países. La caracterización de la epidemia de SIDA y sus factores de riesgo ha dado pauta para el control de bancos de sangre y el fomento de medidas de prevención del riesgo mediante la promoción del uso persistente de condón; asimismo, el estudio de infecciones crónicas y cáncer ha desembocado en la asociación entre el virus del papiloma humano y el cáncer cervical, el virus de la hepatitis y el cáncer de hígado, así como la asociación entre infección por *Helicobacter pylori* y cáncer gástrico. Estos estudios han abierto la posibilidad de llevar a cabo la prevención primaria de estas enfermedades mediante el desarrollo de vacunas profilácticas. La caracterización de patrones de obesidad, dieta y actividad física ha sido ampliamente asociada a la presencia de enfermedades crónicas; el estudio de la distribución y factores asociados a la aparición de lesiones ha logrado que el desarrollo de medidas de prevención y control de los accidentes sea una prioridad actual, y la aplicación de los métodos epidemiológicos para la cuantificación continua de exposición ambiental a ozono, partículas suspendidas en el aire y metales pesados, entre otros, ha sido fundamental para promover ambientes saludables.

Como puede observarse, la aplicación de la epidemiología es diversa y entre sus usos pueden citarse la identificación de necesidades de salud en una comunidad, la caracterización de la historia natural de las enfermedades y la evaluación del impacto de las acciones de salud. La epidemiología no sólo identifica

factores etiológicos, sino que determina posibles mecanismos de transmisión de las enfermedades. Permite, asimismo, predecir el comportamiento de una enfermedad y sus pautas de prevención y control. Recientemente se ha utilizado para probar la eficacia de las estrategias de intervención y la aplicación de sus métodos permite generar insumos que funcionan como evidencias científicas durante la toma de decisiones.

En este contexto, para mí es un gran orgullo presentar a los lectores interesados en la epidemiología una obra que reúne el esfuerzo de un selecto grupo multidisciplinario, que ha aplicado en la práctica cada uno de los diseños de investigación descritos y que ha utilizado una extensa variedad de estrategias de análisis para cada diseño.

Describir a profundidad los métodos utilizados en epidemiología es un privilegio y un ejercicio divertido. Al ser la epidemiología una disciplina científica en permanente evolución, resulta muy difícil delimitarla. Sin embargo, hemos hecho un esfuerzo por describir, de manera clara y práctica, los principales elementos conceptuales y metodológicos que la conforman. Para lograrlo hemos dividido el contenido del libro en dos grandes apartados: el primero se refiere al método propiamente epidemiológico, y en él se hace un recorrido por sus principales diseños; en el segundo se describe la aplicación del análisis estadístico en un nivel de profundidad intermedio, particularmente los modelos estadísticos más utilizados, cuyos procedimientos, en palabras de los autores, buscan "la mejor manera de resumir, analizar y utilizar la información recolectada para responder una pregunta de investigación fraguada en el ámbito de la epidemiología".

El panorama histórico que describe el origen y evolución de los conceptos que paulatinamente fueron dando forma a la epidemiología moderna es tema del primer capítulo. En el segundo se proporciona una explicación general sobre el diseño general de estudios epidemiológicos, con el propósito de orientar al lector acerca de cuál puede ser el más apropiado de acuerdo con los objetivos y alcances de su investigación. Más adelante se dedica un capítulo a cada uno de los diseños considerados más relevantes. El capítulo sobre las principales medidas utilizadas en epidemiología aborda además conceptos que resultan fundamentales para contrastar empíricamente una hipótesis científica.

Los diferentes capítulos dedicados a los diseños de investigación epidemiológica detallan sus premisas básicas, sus alcances y las limitaciones que deben ser consideradas por los investigadores. Así, por ejemplo, al exponer los ensayos aleatorizados se detalla la razón por la cual estos diseños brindan la mayor evidencia posible para confirmar relaciones de causa-efecto entre las intervenciones o exposiciones asignadas aleatoriamente por el investigador; los estudios de cohorte, que son diseños observacionales donde se elige a un grupo de sujetos expuesto y otro no expuesto, a efecto de determinar la ocurrencia del evento, ofrecen también una verificación correcta en el tiempo de la relación causa-efecto; los estudios de casos y controles constituyen una estrategia muestral que trata de simular el desarrollo de un estudio de cohorte. Finalmente se abordan los estudios transversales, en los que no se consideran la exposición ni el efecto como criterios de selección de la muestra, sino que se indagan una vez que ésta ha sido conformada en forma probabilística.

Otros conceptos importantes a los que se han dedicado capítulos enteros son los de investigación de brotes y sesgos epidemiológicos. En el primer caso se describe una metodología capaz de abordar los eventos adversos conocidos como brotes, y ante los cuales el profesional de la epidemiología debe reaccionar rápidamente a fin de identificar los factores asociados a la epidemia, definir medidas que

reduzcan su impacto, recabar información sobre su posible etiología y transmitir las conclusiones así obtenidas de manera oportuna y aplicada a la población.

En cuanto a los sesgos epidemiológicos, se abordan dos nociones fundamentales para garantizar la solidez y pertinencia de los estudios: la validez interna, o ausencia de sesgos en la selección, medición o comparabilidad de los grupos en estudio, y la validez externa, referida a la capacidad de generalizar los resultados a poblaciones, lugares y momentos diferentes a los estudiados.

Los capítulos que se refieren al análisis estadístico parten del supuesto de que el lector conoce las técnicas estadísticas básicas para la exploración de datos, y se aboca por lo tanto a los procedimientos conocidos como modelos estadísticos, usados para analizar la información de manera que responda a los planteamientos de investigación. Algunos de los modelos más importantes que se emplean en la epidemiología se describen en la segunda parte del libro a través de ejemplos, sin la pretensión de ser exhaustivos. Los modelos estadísticos aplicados constituyen una representación de la realidad que se encuentra detrás de las asociaciones encontradas en los datos de investigación. Entre los modelos básicos que se describen se encuentran los de regresión lineal simple, regresión lineal múltiple, regresión logística y análisis de supervivencia.

Agradezco a cada uno de los colaboradores su contribución en la elaboración de esta obra. También es necesario señalar que la mayoría de los capítulos correspondientes a la sección de epidemiología fueron publicados en versiones previas —notablemente distintas de las que ahora presentamos— en la revista *Salud Pública de México*, por lo que agradecemos a la revista su amabilidad al permitirnos reproducir algunas de las figuras, tablas y textos de aquellas primeras comunicaciones.

Por su capacidad de actuar en todos los niveles de la salud pública, la epidemiología es una herramienta fundamental del profesional de la salud. Es, además, una de las ciencias con mayor potencial para generar nuevos conocimientos en el futuro próximo; de ahí la importancia de contar con una oferta mayor de textos en español que permitan al estudioso de la salud profundizar en su conocimiento.

Si la lectura de *Epidemiología. Diseño y análisis de estudios* permite a los interesados manejar con mayor destreza las herramientas metodológicas necesarias para generar información útil en la toma de decisiones basadas en la evidencia científica, habremos alcanzado nuestro propósito.

MAURICIO HERNÁNDEZ ÁVILA

# I Desarrollo histórico de la epidemiología

*Sergio López Moreno*
*Mauricio Hernández Ávila*

La epidemiología es la ciencia que tiene como propósito describir y explicar la dinámica de la salud poblacional, identificar los elementos que la componen y comprender las fuerzas que la gobiernan, a fin de desarrollar acciones tendientes a conservar y promover la salud de la población. La epidemiología investiga la distribución, frecuencia y determinantes de las condiciones de salud en las poblaciones humanas así como las modalidades y el impacto de las respuestas sociales instauradas para atenderlas. La epidemiología también se encarga de desarrollar los métodos y técnicas necesarios para cumplir con sus objetivos.

Para la epidemiología, el término condiciones de salud no se limita a la ocurrencia de enfermedades y, por esta razón, su estudio incluye todos aquellos eventos relacionados directa o indirectamente con la salud, comprendiendo este concepto en forma amplia. En consecuencia, la epidemiología investiga, bajo una perspectiva poblacional: a) la distribución, frecuencia y determinantes de la enfermedad y sus consecuencias biológicas, psicológicas y sociales; b) la distribución y frecuencia de los marcadores de enfermedad; c) la distribución, frecuencia y determinantes de los riesgos para la salud; d) las formas de control de las enfermedades, de sus consecuencias y sus riesgos; e) las modalidades e impacto de las respuestas adoptadas para atender todos estos eventos, y f) los métodos necesarios para mejor comprender las dimensiones poblacionales de la enfermedad. Para su desarrollo, en la práctica de la epidemiología se combinan principios y conocimientos generados por las ciencias biológicas y sociales, y se aplican metodologías de naturaleza cuantitativa y cualitativa.

La transformación de la epidemiología en una ciencia ha tomado varios siglos, y puede decirse que es una ciencia joven. Todavía en 1928, el epidemiólogo inglés Clifford Allchin Gill[1] señalaba que la disciplina, a pesar de su antiguo linaje, se encontraba en la infancia. Como muestra, afirmaba que los escasos logros obtenidos por ella en los últimos 50 años no le permitían reclamar un lugar entre las ciencias exactas; que apenas si tenía alguna literatura especializada, y que en vano podían buscarse sus libros de texto; dudaba incluso que los problemas que abordaba estuviesen claramente comprendidos por los propios epidemiólogos. Siete décadas después, el panorama descrito por Gill es ciertamente diferente, existen numerosos libros de texto y actualmente pocos son los avances en la medicina o políticas públicas relacionadas con la salud que no cuenten con la participación de conocimiento generado a través del método epidemiológico. En la actualidad se reconoce a la epidemiología como una de las ciencias troncales de la salud pública.

## PLAGAS Y EPIDEMIAS

La reflexión sobre las enfermedades como fenómenos colectivos es casi tan antigua como la escritura, y las primeras descripciones de padecimientos que afectan a poblaciones enteras posiblemente se refieren a enfermedades de naturaleza infecciosa. El papiro de Ebers, que menciona unas fiebres pestilentes —probablemente malaria— que asolaron a la población de las márgenes del Nilo alrededor del año 1500 a.C., es probablemente el texto en el que se hace la referencia más antigua a un padecimiento colectivo.[2] La aparición periódica de plagas a partir del momento en la historia en el que los grupos adquirieron cierta densidad poblacional parece indiscutible. En Egipto, hace 3000 años, se veneraba a una diosa, llamada *Sekmeth*, que se suponía capaz de provocar la peste; también existen momias de entre dos mil y tres mil años de antigüedad que muestran afecciones que sugieren infecciones trasmisibles, como la poliomielitis.[3-5] Dado que la momificación estaba reservada a los personajes más importantes del antiguo Egipto —que posiblemente no tenían contacto con grandes grupos poblacionales, excepto a través de los esclavos—, no sería extraño que este tipo de afecciones fuera mucho más frecuente entre la población general.

La aparición de plagas a lo largo de la historia también fue registrada en la mayor parte de los libros sagrados, en especial en la Biblia, el Talmud y el Corán, que adicionalmente contienen las primeras normas para prevenir las enfermedades contagiosas. De estas descripciones destaca la de la plaga que obligó a Mineptah, el faraón egipcio que sucedió a Ramsés II, a permitir la salida de los judíos de Egipto, alrededor del año 1224 a.C.[6]

Muchos escritores griegos y latinos se refirieron a menudo al surgimiento de lo que denominaron pestilencias. La más famosa de estas descripciones es quizás la de la plaga de Atenas, que asoló esta ciudad durante la Guerra del Peloponeso entre los años 430-427 a. C. y que según Tucídides se inicia en Etiopía, pasa por Egipto y los Pireos llegando finalmente a Atenas.* Antes y después de este historiador, otros escritores occidentales como Homero, Herodoto, Lucrecio, Ovidio y Virgilio[7-9] se refieren al desarrollo de procesos morbosos colectivos que, sin duda, pueden considerarse fenómenos epidémicos. Una de las características más notables de estas descripciones es que dejan muy claro que la mayoría de la población creía firmemente que muchos padecimientos eran contagiosos. Las acciones preventivas y de control de las afecciones contagiosas también son referidas en muchos textos antiguos. Como ya hemos dicho, la Biblia, el Corán, el Talmud y diversos libros chinos e hindúes recomiendan numerosas prácticas sanitarias preventivas, como el lavado de manos y alimentos, la circuncisión, el aislamiento de enfermos, la inhumación o cremación de los cadáveres y el uso del condón. Por los Evangelios sabemos que algunos enfermos contagiosos —como los leprosos— eran invariablemente aislados y tenían prohibido establecer comunicación con la población sana. Por su parte, el uso del condón puede documentarse en épocas tan antiguas como 1000 a.C., pues se sabe que los antiguos egipcios utilizaban fundas de lino probablemente como protección contra enfermedades. Los condones más antiguos documentados corresponden al 1640 y se encontraron en excavaciones realizadas en Birmingham, en Inglaterra. En el siglo XVI, Gabriele Fallopius publicó la primera descripción del condón como medida preventiva. Fallopius también fue el pri-

* Jarcho S. The concept of contagion in Medicine, literature and religion. Krieger Publishing Company, 2000. Malabar, Florida, USA.

mero en llevar a cabo ensayos clínicos sobre la efectividad del preservativo. Aseguró haber inventado un preservativo elaborado con hilo que ensayó en 1 100 hombres sin que ninguno de ellos contrajera la sífilis.*

La palabra epidemiología, que proviene de los términos griegos "epi" (encima), "demos" (pueblo) y "logos" (estudio), etimológicamente significa el estudio de "lo que está sobre las poblaciones". La primera referencia propiamente médica de un término análogo se encuentra en Hipócrates (460-385 a.C.), quien usó las expresiones *epidémico* y *endémico* para referirse a los padecimientos según fueran o no propios de determinado lugar.[10] Aunque los textos Hipocráticos describen con detalle la ocurrencia de enfermedades en grupos poblacionales, así como sus síntomas, complicaciones y secuelas, en ellos no se menciona o reconoce la posibilidad de contagio de persona a persona. Estos textos atribuyen la aparición de las enfermedades al ambiente malsano (*miasmas*) y a la falta de moderación en la dieta y las actividades físicas. El libro conocido como *Aires, aguas, y lugares* —que sigue la teoría de los elementos propuesta medio siglo antes por el filósofo y médico Empédocles de Agrigento— señala que la dieta, el clima y la calidad de la tierra, los vientos y el agua son los factores involucrados en el desarrollo de las enfermedades en la población, al influir sobre el equilibrio del hombre con su ambiente. Siguiendo estos criterios, Hipócrates elabora el concepto de *constitución epidémica* de las poblaciones.

Aunque la noción de balance entre el hombre y su ambiente como sinónimo de salud persistió por muchos siglos, con el colapso de la civilización clásica el Occidente retornó a las concepciones mágico-religiosas que caracterizaron a las primeras civilizaciones.[11] Con ello, la creencia en el contagio como fuente de enfermedad, común a casi todos los pueblos antiguos, paulatinamente fue subsumida por una imagen en donde la enfermedad y la salud significaban el castigo y el perdón divinos, y las explicaciones sobre la causa de los padecimientos colectivos estuvieron prácticamente ausentes en los escritos médicos elaborados entre los siglos III y XV de nuestra era (es decir, durante el periodo en el que la Iglesia Católica gozó de una hegemonía casi absoluta en el terreno de las ciencias). No obstante, como veremos más tarde, las medidas empíricas de control de las infecciones epidémicas siguieron desarrollándose, gracias a su impacto práctico.

Durante el reinado del emperador Justiniano, entre los siglos V y VI d. C., la terrible plaga que azotó el mundo ya recibió el nombre griego de "epidemia". No se sabe exactamente desde cuándo el término *epidémico* se usa para referirse a la presentación de un número inesperado de casos de enfermedad, pero no hay duda de que el término fue utilizado desde la baja Edad Media para describir el comportamiento de las infecciones que, periódicamente, devastaban a las poblaciones. La larga historia de epidemias infecciosas que azotaron el mundo antiguo y medieval fue determinando una identificación casi natural entre los conceptos de epidemia, infección y contagio hasta que, según Winslow, la aparición de la pandemia de peste bubónica o peste negra que azotó Europa durante el siglo XIV (de la cual se dice que diariamente morían 10 mil personas), finalmente condujo a la aceptación universal —aunque todavía en el ámbito popular— de la doctrina del contagio.[7]

Los esfuerzos por comprender la naturaleza de las enfermedades y su desarrollo entre la población condujeron a la elaboración de diversas obras médicas durante los siglos inmediatamente posteriores al Renacimiento. En 1546, Girolamo Fracastoro (1478-1553) publicó, en Venecia, el libro *De contagione et con-*

* Youssef, H. 1993. "The history of the condom." J Royal Soc Med 86:266-228

*tagiosis morbis et eorum curatione*, en donde por primera vez describe todas las enfermedades que en ese momento podían calificarse como contagiosas (peste, lepra, tisis, sarna, rabia, erisipela, viruela, ántrax y tracoma) y agrega, como entidades nuevas, el tifus exantemático y la sífilis. Fracastoro fue el primero en establecer claramente el concepto de contagio *(quaedam abuno ad aliud transiens infectio)* y propone sus tres formasx posibles: a) por contacto directo *(seminaria contagionum)*, mediante semillas vivas capaces de provocar la enfermedad; b) por medio de fomites (*fomes)*, que son los conductores de los *seminaria prima*, y c) a distancia o por medio del aire *(mistio)*. A este médico italiano también le cabe el honor de establecer en forma precisa la separación, actualmente tan clara, entre los conceptos de infección (causa) y epidemia (consecuencia). Como veremos más adelante, incluso para médicos tan extraordinarios como Thomas Sydenham —quien nació cien años más tarde que Fracastoro y popularizó el concepto hipocrático de *constituciones epidémicas*, y los de higiene individual y poblacional de Galeno— fue imposible comprender esta diferencia fundamental. Fracastoro también fue el primero en establecer que enfermedades específicas resultan de contagios específicos, presentando la primera teoría general del contagio vivo de la enfermedad. Desde este punto de vista, debe ser considerado el padre de la epidemiología moderna.[12]

Treinta y cuatro años después de Fracastoro, en 1580, el médico francés Guillaume de Baillou (1538-1616) publicó el libro *Epidemiorum* ("sobre las epidemias") que contenía una relación completa de las epidemias de sarampión, difteria y peste bubónica, aparecidas en Europa entre 1570 y 1579, así como sus características y modos de propagación. Debido a que de Baillou tuvo una gran influencia en la enseñanza de la medicina durante la última parte del siglo XVI y la primera del XVII (dirigió la escuela de medicina de la Universidad de París durante varias décadas), sus trabajos tuvieron un importante impacto en la práctica médica de todo el siglo XVII.

En castellano, la primera referencia al término epidemiología, según Nájera,[13] se encuentra en el libro que con tal título publicó Quinto Tiberio Angelerio, en Madrid, en 1598. Los términos epidémico y endémico fueron incorporados a nuestra lengua apenas unos años más tarde, hacia 1606. En 1732, cuando aparece el primer diccionario de la lengua española, llamado comúnmente Diccionario de Autoridades, el término "epidemia" es una de primeras palabras incluidas. Debe señalarse que en aquella época la palabra endémico significaba simplemente (como en el texto hipocrático *Aires, aguas y lugares*) aquello que reside en el lugar de donde es originario. Epidémico, en cambio, designaba a quien temporalmente reside en un lugar donde es extranjero.[14]

Desde mucho antes, empero, el Occidente medieval había llevado a cabo actividades de salud pública que podrían calificarse como epidemiológicas en el sentido actual del término. La Iglesia ejecutó durante muchos siglos acciones de control sanitario destinadas a mantener lejos del cuerpo social las enfermedades que viajaban con los ejércitos y el comercio, y tempranamente aparecieron prácticas sanitarias que basaban su fuerza en los resultados del aislamiento y la cuarentena. Desde el siglo XIV hasta el XVII, estas acciones se generalizaron en toda Europa y paulatinamente se incorporaron a la esfera médica.

## LA APARICIÓN DE LA ESTADÍSTICA SANITARIA

Durante los siguientes siglos ocurrieron en Europa otros sucesos de naturaleza diferente que, sin embargo, tuvieron un fuerte impacto sobre el desarrollo de la epidemiología. Hasta el siglo XVI, la mayoría de las enumeraciones y recuentos poblacionales habían tenido casi exclusivamente dos propósitos: determinar la carga de impuestos y reclutar miembros para el ejército. No obstante, con el nacimiento de las naciones modernas, los esfuerzos por conocer de manera precisa las fuerzas del Estado (actividad que inicialmente se denominó *estadística*) terminaron por rebasar estos límites e inaugurar la cuantificación sistemática de un sinnúmero de características entre los habitantes de las florecientes naciones europeas. La estadística de salud moderna inició con el análisis de los registros de nacimiento y mortalidad, hasta entonces realizados únicamente por la Iglesia Católica, que organizaba sus templos de acuerdo con el volumen de sus feligreses.

El nacimiento de las estadísticas sanitarias coincide con un extraordinario avance de las ciencias naturales (que en ese momento hacían grandes esfuerzos por encontrar un sistema lógico de clasificación botánica) y que se reflejó en las cuidadosas descripciones clínicas de la disentería, malaria, viruela, gota, sífilis y tuberculosis, hechas por el inglés Thomas Sydenham entre 1650 y 1676. Los trabajos de este autor resultaron esenciales para reconocer estas patologías como entidades distintas y dieron origen al sistema actual de clasificación de enfermedades. En su libro *Observationes medicae*, Sydenham afirmaba, por ejemplo, que si la mayoría de las enfermedades podían agruparse de acuerdo con los criterios de "unidad biológica", también era posible reducirlas a unos cuantos tipos, "exactamente como hacen los botánicos en sus libros sobre las plantas".[15] Las propuestas clasificatorias abiertas por Sydenham se vieron fortalecidas casi inmediatamente, cuando su coterráneo John Graunt analizó, en 1662, los reportes semanales de nacimientos y muertes observados en la ciudad de Londres y el poblado de Hampshire durante los 59 años previos, e identificó un patrón constante en las causas de muerte y diferencias entre las zonas rurales y urbanas.[12] John Graunt fue un hombre extraordinariamente perspicaz. Disponiendo de información mínima logró inferir, entre otras cosas, que regularmente nacían más hombres que mujeres; que había una clara variación estacional en la ocurrencia de las muertes, y que podía predecirse que 36% de los nacidos vivos morirían antes de cumplir los seis años. Con ello, Graunt dio los primeros pasos para el desarrollo de las actuales tablas de vida y, en consecuencia, de la demografía y epidemiología modernas.

Un economista, músico y médico amigo de Graunt, William Petty, publicó por la misma época varios trabajos relacionados con los patrones de mortalidad, natalidad y enfermedad entre la población inglesa, y propuso por primera vez —30 años antes que Leibniz (1646-1716), a quien tradicionalmente se le atribuye esta idea— la creación de una agencia gubernamental encargada de la recolección e interpretación sistemática de la información sobre nacimientos, matrimonios y muertes, y su distribución según sexo, edad, ocupación, nivel educativo y otras condiciones de vida. También sugirió la construcción de tablas de mortalidad por edad de ocurrencia, con lo que se anticipó al desarrollo de las actuales tablas utilizadas para comparar poblaciones diferentes. Esta manera de tratar la información poblacional fue denominada por Petty "política aritmética".[15] Los trabajos de Graunt y Petty no contribuyeron inmediatamente a la comprensión de la naturaleza de la enfermedad, pero fueron fundamentales para establecer los sistemas de recolección y organización de la información

que actualmente constituyen las bases de la vigilancia epidemiológica.

En los siguientes años, el estudio de la enfermedad poblacional bajo este método condujo a la elaboración de un sinnúmero de "leyes de la enfermedad", que inicialmente se referían a la probabilidad de enfermar a determinada edad; de permanecer enfermo durante un tiempo específico, y de fallecer por determinadas causas de enfermedad. Estos cálculos, sin embargo, no fueron desarrollados por científicos como Graunt y Petty, sino por las compañías aseguradoras que con gran afán buscaban fijar adecuadamente los precios de los seguros de vida, comunes en Inglaterra y Gales desde mediados del siglo XVII, y en Francia desde mucho tiempo atrás (quizás desde el siglo XVI) a través de las asociaciones de socorro mutuo y las "tontinas" de trabajadores. Entre los más famosos constructores de tablas de vida para las compañías aseguradoras se encuentran el astrónomo británico Edmund Halley (1656-1742), descubridor del cometa que lleva su nombre y que en 1687 sufragara los gastos de publicación de los *Principia mathematica* de su amigo Isaac Newton, y el periodista Daniel Defoe (1660-1731), autor de la novela *Robinson Crusoe* y del extraordinario relato sobre la epidemia londinense de 1665, *Diario del año de la peste*. Las más famosas tablas elaboradas para estos fines fueron las de los *comités seleccionados*, en Suecia; las de Richard Price, en Inglaterra, y las de Charles Oliphant (siglo XIX), en Escocia. Las más exactas, elaboradas por Price,[16] permiten saber que el promedio de vida en la ciudad de Northampton en el siglo XVIII era de 24 años.

El proceso matemático que condujo a la elaboración de "leyes de la enfermedad" entre los científicos inició con el análisis de la distribución de los nacimientos. En 1710, John Arbuthnot, continuador de los trabajos de Graunt y Petty, había demostrado que la razón de nacimientos entre varones y mujeres era siempre de 13 a 12, independientemente de la sociedad y el país en el que se estudiaran. Para Arbuthnot, esta regularidad no podía deberse al azar, y tenía que ser una "disposición divina" encaminada a balancear el exceso de muertes masculinas debidas a la violencia y la guerra.[16] Entre 1741 y 1775, el sacerdote alemán J. P. Sussmilch escribió varios tratados que seguían los métodos de enumeración propuestos por Graunt, Petty y Arbuthnot. Para Sussmilch, la regularidad encontrada en el volumen de nacimientos por sexo era toda una "ley estadística" (como las leyes naturales de la física) y debían existir leyes similares capaces de explicar el desarrollo de toda la sociedad. Muy pronto nació la idea de una "ley de mortalidad" y, poco más tarde, la convicción de que debería haber leyes para todas las desviaciones sociales: el suicidio, el crimen, la vagancia, la locura y, naturalmente, la enfermedad.[16] Si bien las estadísticas sobre la enfermedad tuvieron importancia práctica sólo hasta el siglo XIX, su desarrollo fue un avance formidable para la época. La misma frase "ley de la enfermedad" invitaba a formular los problemas de salud en forma matemática, generalizando estudios sobre la causa de los padecimientos y muertes entre la población. En 1765, el astrónomo Johann H. Lambert inició la búsqueda de relaciones entre la mortalidad, el volumen de nacimientos, el número de casamientos y la duración de la vida, usando la información de las gacetas estadísticas alemanas. Como resultado, Lambert obtuvo una curva de decesos que incorporaba la duración de vida promedio de la población investigada y con la cual logró deducir una tasa de mortalidad infantil mucho más alta de lo que entonces se pensaba.

La búsqueda de "leyes de la enfermedad" fue una actividad permanente hasta el final del siglo XIX, y contribuyó al desarrollo de la estadística moderna.[17] Durante este proceso, la incorporación de la probabilidad en el estudio de la enfermedad fue un proceso casi natural.

## LA "OBSERVACIÓN NUMÉRICA" Y LA COMPRENSIÓN DE LAS CAUSAS DE ENFERMEDAD

Para la misma época, por otra parte, se habían publicado trabajos que también hacían uso, aunque de otra manera, de la enumeración estadística. El primero de ellos, publicado en 1747, fue un trabajo de James Lind sobre la etiología del escorbuto, en el que demostró experimentalmente que la causa de esta enfermedad era un deficiente consumo de cítricos. El segundo fue un trabajo publicado en 1760 por Daniel Bernoulli, que concluía que la variolación natural protegía contra la viruela y confería inmunidad de por vida.[12] Es notable que este trabajo se publicara 38 años antes de la introducción del método de vacunación propuesto por el británico Edward Jenner (1749-1823). Un tercer trabajo, que se refiere específicamente a la práctica de inmunización introducido por Jenner, fue publicado por Duvillard de Durand apenas nueve años después de la generalización de este procedimiento en Europa (en 1807), y se refiere a las potenciales consecuencias de este método preventivo en la longevidad y la esperanza de vida de los franceses.[16]

No obstante, como señala Hacking, el imperialismo de las probabilidades sólo era concebible en un *mundo numérico*. Aunque la cuantificación se hizo común a partir de Galileo, en materia médica, esto fue posible sólo gracias a los trabajos de Pierre Charles Alexander Louis. Este clínico francés, uno de los primeros epidemiólogos modernos, condujo, a partir de 1830, una gran cantidad de estudios de observación "numérica", demostrando, entre muchas otras cosas, que la tuberculosis no se trasmitía hereditariamente y que la sangría era inútil y aun perjudicial en la mayoría de los casos.[16] La enorme influencia de Pierre Charles Alexander Louis durante las siguientes décadas se muestra en la primera declaración de la Sociedad Epidemiológica de Londres, fundada en 1850, en la que se afirma que "la estadística también nos ha proporcionado un medio nuevo y poderoso para poner a prueba las verdades médicas, y mediante los trabajos del preciso Louis hemos aprendido cómo puede ser utilizada apropiadamente para entender lo relativo a las enfermedades epidémicas".

El mayor representante de los estudios sobre la regularidad estadística en el siglo XIX fue, sin embargo, el belga Adolphe Quetelet, quien usó los estudios de Poisson y Laplace para identificar los valores promedio de múltiples fenómenos biológicos y sociales. Como resultado, Quetelet transformó cantidades físicas conocidas en propiedades ideales que seguían comportamientos regulares, con lo que inauguró los conceptos de término medio y normalidad biológica, categorías ampliamente usadas durante la inferencia epidemiológica. Sin embargo, los trabajos de Laplace, Louis, Poisson, Quetelet, Galton y Pearson pronto se acercaron a las posturas sostenidas por los científicos positivistas (especialmente los físicos), para quienes, según el dicho del escocés William Kelvin, una ciencia que no medía "era una pobre ciencia". Con ello, se pasó de considerar que *medir es bueno*, a creer que *sólo medir es bueno*.

Un alumno distinguido de Louis, el inglés William Farr, generalizó el uso de las tasas de mortalidad y también los conceptos de población bajo riesgo, gradiente dosis-respuesta, inmunidad de grupo, direccionalidad de los estudios y valor "año-persona". También descubrió las relaciones entre prevalencia, incidencia y duración de las enfermedades, y fundamentó la necesidad de contar con grandes grupos de casos para lograr inferencias válidas.[12] En 1837 publicó lo que denominó "un instrumento capaz de medir la frecuencia y duración relativa de las enfermedades", afirmando que con él era

posible determinar el *peligro relativo* de cada padecimiento. Finalmente, creó el concepto de *fuerza de la mortalidad* de un padecimiento específico, y lo definió como el volumen de "decesos entre un número determinado de enfermos del mismo padecimiento, en un periodo definido de tiempo".[16] Este, que es uno de los primeros conceptos epidemiológicos altamente precisos, es idéntico al que hoy conocemos como letalidad.

La investigación realizada en el campo de la epidemiología experimentó durante el siglo XIX un extraordinario avance, especialmente con los trabajos de Robert Storrs (1840), Oliver Wendell Holmes (1842) e Ignaz Semmelweis (1848) sobre la transmisión de la fiebre puerperal; los de P. L. Panum (1846) sobre la contagiosidad del sarampión; los de Snow (1854) sobre el modo de transmisión del cólera, y los de William Budd (1857) sobre la transmisión de la fiebre tifoidea. En América Latina destacan los trabajos realizados por Carlos Finlay sobre el papel del mosquito en la transmisión de la fiebre amarilla, los de Daniel Carrión sobre la fiebre de Oroya y, ya en pleno siglo XX, los de Carlos Chagas y Oswaldo Cruz en la investigación del agente etiológico de la tripanosomiasis.

La importancia de estos trabajos radica en el enorme esfuerzo intelectual que estos investigadores debieron hacer para documentar, mediante la pura observación —naturalmente, observación guiada por la teoría—, propuestas sobre la capacidad transmisora, los mecanismos de contagio y la infectividad de agentes patógenos sobre los que aún no podía demostrarse una existencia real. Una muestra del enorme valor de este trabajo se encuentra en el hecho de que los agentes infecciosos responsables de cada una de estas enfermedades se descubrieron entre veinte y treinta años más tarde, en el mejor de los casos.

El método utilizado por los epidemiólogos del siglo XIX para demostrar la transmisibilidad y contagiosidad de los padecimientos mencionados (que, en resumen, consiste en comparar, de múltiples formas, la proporción de enfermos expuestos a una circunstancia con la proporción de enfermos no expuestos a ella) se reprodujo de manera sorprendente, y con él se estudiaron, durante los siguientes años, prácticamente todos los brotes epidémicos. De hecho, versiones más sofisticadas de esta estrategia constituyen actualmente los principales métodos de la epidemiología.

La escuela de epidemiólogos fundada en el siglo pasado continúa activa. Las ideas de Louis, por ejemplo, fueron adoptadas por muchos de sus alumnos y siguen dando frutos. Entre sus alumnos destacan Francis Galton (descubridor del coeficiente de correlación), George C. Shattuck (fundador de la Asociación Estadística Norteamericana y reformador de la salud pública en ese país) y Elisha Bartlett (el primero en justificar matemáticamente el uso del *grupo control* en los estudios experimentales). Un alumno de Galton, Karl Pearson, descubrió la distribución de $\chi^2$ y fundó la Escuela Británica de Biometría. Major Greenwood, alumno de Pearson, fue el más destacado epidemiólogo inglés de la primera mitad del siglo XX y maestro de Austin Bradford Hill quien, junto con Evans y Jerushalmy, ha sido uno de los más importantes divulgadores de los criterios modernos de causalidad. En nuestro continente destacaron inicialmente Edward Jarvis, William Welch, Joseph Goldberger, Wade Hampton Frost, Edgard Sydenstriker y Kenneth Maxcy. Más recientemente, ambas escuelas epidemiológicas han dado nombres de la talla de Richard Doll, Jerome Cornfield, Alexander Langmuir, Brian MacMahon, Nathan Mantel, William Haenzel, Abraham Lilienfeld, Thomas Mckeown, Milton Terris, Carol Buck, Mervyn Susser, Sanders Greenland, Olli Miettinen, David Kleimbaum y Kenneth Rothman, todos ellos reconocidos por sus importantes contribuciones al desarrollo metodológico de la disciplina.

## DISTRIBUCIÓN Y FRECUENCIA DE LAS CONDICIONES DE SALUD

Con el establecimiento definitivo de la teoría del germen, entre 1872 y 1880, la epidemiología, como todas las ciencias de la salud, adoptó un modelo de causalidad que reproducía el de la física, y en el que un solo efecto es resultado de una sola causa, siguiendo conexiones lineales. Los seguidores de esta teoría fueron tan exitosos en la identificación de la etiología específica de enfermedades que dieron gran credibilidad a este modelo. Como consecuencia, la epidemiología comenzó a utilizarse de manera muy importante en el estudio de las enfermedades infecciosas.

Las experiencias de investigación posteriores rompieron estas restricciones. Las realizadas entre 1914 y 1923 por Joseph Goldberger —quien demostró el carácter no contagioso de la pelagra— rebasaron los límites de la infectología y sirvieron de base para elaborar teorías y adoptar medidas preventivas eficaces contra las enfermedades carenciales, inclusive antes de que se conociera el modo de acción de los micronutrimentos esenciales.[13] En 1936, Frost afirmaba que la epidemiología "en mayor o menor grado, sobrepasa los límites de la observación directa", asignándole la posibilidad de un desarrollo teórico propio y, en 1941, Major Greenwood la definió simplemente como "el estudio de la enfermedad, considerada como fenómeno de masas".

El incremento en la incidencia de enfermedades crónicas ocurrido a mediados del siglo XX también contribuyó a ampliar el campo de acción de la disciplina, la que desde los años cuarenta se ocupó del estudio de la dinámica del cáncer, la hipertensión arterial, las afecciones cardiovasculares, las lesiones y los padecimientos mentales y degenerativos. Como resultado, la epidemiología desarrolló con mayor precisión los conceptos de exposición, riesgo, asociación, confusión y sesgo, e incorporó el uso franco de la teoría de la probabilidad y de un sinnúmero de técnicas de estadística avanzada.[18]

## EL PARADIGMA DE LA RED CAUSAL

Desde su nacimiento como disciplina moderna, una premisa fundamental de la epidemiología ha sido la afirmación de que la enfermedad no ocurre ni se distribuye al azar, y sus investigaciones tienen como propósito identificar claramente las condiciones que pueden ser calificadas como "causas" de las enfermedades, distinguiéndolas de las que se asocian a ellas únicamente por azar.[19,20] El incesante descubrimiento de condiciones asociadas a los procesos patológicos ha llevado a la identificación de una intrincada red de "causas" para cada padecimiento, y desde los años setenta se postula que el peso de cada factor presuntamente causal depende de la cercanía lógica, espacial y temporal con su efecto aparente. La epidemiología contemporánea ha basado sus principales acciones en este modelo, denominado "red de causalidad" y formalizado por Brian MacMahon, en 1970.

Una versión más acabada de este mismo modelo propone que las relaciones establecidas entre las condiciones participantes en el proceso —denominadas causas, o efectos, según su lugar en la red— son tan complejas, que forman una unidad imposible de conocer completamente. El modelo, conocido como de la "caja negra", es la metáfora con la que se representa un fenómeno cuyos procesos internos están ocultos al observador, y sugiere que

la epidemiología debe limitarse a la búsqueda de aquellas partes de la red en las que es posible intervenir efectivamente, rompiendo la cadena causal y haciendo innecesario conocer todos los factores intervinientes en el origen de la enfermedad. Actualmente, este es el modelo predominante en la investigación epidemiológica.[21,22] Una de sus principales ventajas radica en la posibilidad de aplicar medidas correctivas eficaces, aun sin dilucidar completamente la cadena de causalidad o los mecanismos íntimos que desencadenan los eventos. Esto sucedió, por ejemplo, cuando en la década de los cincuenta se identificó la asociación entre el cáncer pulmonar y el hábito de fumar.[23] No era necesario conocer los mecanismos cancerígenos precisos de inducción y promoción para abatir la mortalidad mediante el combate al tabaquismo. Una desventaja del modelo, empero, es que con frecuencia existe una deficiente comprensión de los eventos que se investigan, al no ser necesario comprender todo el proceso para adoptar medidas eficaces de control. El resultado más grave del seguimiento mecánico de este esquema ha consistido en la búsqueda desenfrenada de "factores de riesgo" sin esquemas explicativos sólidos, lo que ha hecho parecer a los estudios epidemiológicos como una colección infinita de factores que, en última instancia, explican muy poco los orígenes de las enfermedades.

El modelo de la caja negra también tiene como limitación la dificultad para distinguir entre los determinantes individuales y poblacionales de la enfermedad (es decir, entre las causas de los casos y las causas de la incidencia). Geoffrey Rose ha advertido esta falta de discriminación al preguntarse si la aparición de la enfermedad en las personas puede explicarse de la misma manera que la aparición de la enfermedad en las poblaciones.[24] En otras palabras, Rose se pregunta si la enfermedad individual y la incidencia tienen las mismas causas y, por lo tanto, si pueden combatirse con las mismas estrategias. Rose mismo responde esta pregunta diciendo que no.

Corrientes más recientes han intentado desarrollar un paradigma opuesto al de la caja negra multicausal, denominado modelo histórico-social. Este modelo señala que es engañoso aplicar mecánicamente un modelo que concede el mismo peso a factores que, por su naturaleza, deben ser diferentes. También rechaza que el componente biológico de los procesos de salud colectiva tenga un carácter determinante, y propone reexaminar estos fenómenos a la luz de su determinación histórica, económica y política. Según esta interpretación, el propósito principal de la investigación epidemiológica debe ser la explicación de la distribución desigual de las enfermedades entre las diversas clases sociales, en donde se encuentra la determinación de la salud-enfermedad.[25] No obstante el interés que revisten estos planteamientos, el limitado desarrollo de instrumentos conceptuales adecuados para contrastar sus hipótesis ha impedido que este modelo progrese como una alternativa real a los modelos de la red de causalidad y de la caja negra.



## LA ECOEPIDEMIOLOGÍA

Entre los trabajos que directamente abordan el problema de la "caja negra" destaca la obra de Mervyn Susser.[26] Para él, los fenómenos colectivos de salud funcionan de manera más parecida a una "caja china", en donde los sistemas de determinación epidemiológica se encuentran separados y organizados jerárquicamente, de forma tal que un sistema abarca varios subsistemas, compuestos a su vez por subsistemas de menor jerarquía. Así, los cambios en

un nivel afectan al subsistema correspondiente, pero nunca al sistema en su totalidad. De esta manera, las relaciones de cada nivel son válidas para explicar estructuras en los nichos de donde se han obtenido; pero no para realizar generalizaciones en otros niveles. Esta propuesta, denominada ecoepidemiología, explica, por ejemplo, la razón por la cual la información obtenida en el subsistema donde se enmarca y determina la desnutrición biológica individual no puede explicar los sistemas en los que se enmarcan y determinan la incidencia de desnutrición de una comunidad, una región o un país.

## FUNCIONES FUNDAMENTALES DE LA EPIDEMIOLOGÍA MODERNA

### Determinación de riesgos

Como anteriormente sucedió con las enfermedades infecciosas, en el estudio de las afecciones crónicas y degenerativas, la epidemiología ha vuelto a jugar un papel fundamental, al mostrar la relación que existe entre determinadas condiciones del medio ambiente físico o social, estilo de vida y carga genética, y la aparición de daños específicos en las poblaciones en riesgo. Entre sus aportes más importantes se encuentran, por ejemplo, la comprobación de la relación existente entre el consumo de cigarrillos y el cáncer de pulmón; entre la contaminación ambiental y la mortalidad prematura; entre la exposición a plomo y la disminución del coeficiente intelectual; entre la exposición a radiaciones ionizantes y determinadas formas de cáncer; entre la exposición a diversas sustancias químicas o minerales como el asbesto y el incremento de tumores malignos; entre la obesidad y la diabetes mellitus; entre el consumo de estrógenos y cáncer endometrial; entre uso de fármacos y las malformaciones congénitas, y entre sedentarismo y el infarto de miocardio. En la década de los ochenta, diversos estudios epidemiológicos encontraron una fuerte asociación entre las prácticas sexuales y el riesgo de transmisión del Síndrome de Inmunodeficiencia Humana, aun antes del descubrimiento del virus responsable de su aparición. Como antes lo hizo para los padecimientos infecciosos y las enfermedades carenciales, la investigación epidemiológica sigue jugando un extraordinario papel en la identificación de nuevos riesgos, y abre caminos para la toma de medidas preventivas selectivas entre las poblaciones en riesgo.

### Identificación de marcadores de enfermedad

El campo de acción de la epidemiología se amplía permanentemente. Con el surgimiento de la genética y la biología molecular, los epidemiólogos han podido plantearse y responder nuevas preguntas. Ahora se investiga con métodos epidemiológicos, por ejemplo, la distribución poblacional de genes que podrían explicar las variaciones en la presentación de diversos padecimientos neoplásicos, muchas enfermedades endocrinas y no pocas enfermedades mentales y neurológicas. De hecho, la epidemiología se encuentra en franco intercambio teórico y metodológico con la genética y la biología molecular, y su contribución al entendimiento del proceso causal de un número cada vez mayor de padecimientos es creciente, como se ha señalado recientemente:

> Hasta qué punto las promesas de la genética se van a cumplir en un futuro próximo, es materia de debate. [No obstante,] aunque hasta ahora estas promesas parecen ser exageradas o, cuando menos, prematuras, no hay duda de que se esperan

> muchos avances y nuevas formas de unión entre la epidemiología genética y la investigación médica [...] Afortunadamente, todas las discusiones apuntan hacia la mejor comprensión de la compleja relación existente entre los genes y el ambiente, y paulatinamente se han abandonado las posturas que proponían la preeminencia de unos sobre el otro, o viceversa. En esta búsqueda, la epidemiología tendrá un papel determinante, como lo tuvo en la comprensión de las relaciones entre la salud individual y las condiciones colectivas de salud.[27]

**Comprensión de la dinámica general de la enfermedad**

La identificación del comportamiento epidemiológico de los padecimientos según la edad, el género y la región que afectan ha contribuido a la elaboración de teorías generales sobre la dinámica espacial y temporal de la enfermedad, considerada como un fenómeno natural y social. Actualmente, por ejemplo, ya nadie niega que a cada tipo de sociedad corresponde un perfil específico de enfermedad, y que este perfil está ligado al volumen y la estructura de su población, su organización socioeconómica y su capacidad para atender la enfermedad. En este caso, la epidemiología ha representado un papel protagónico al identificar las fases del cambio sanitario y los mecanismos a partir de los cuales un grupo de patologías, característico de una sociedad determinada, es sustituido por otro, propio de una nueva fase. De acuerdo con la teoría de la transición epidemiológica, todos los países deben atravesar tres grandes eras, y la mayoría se encuentra en transición entre la segunda y la tercera fase del proceso. Siguiendo esta teoría, las enfermedades se han reclasificado según el sitio que teóricamente deberían ocupar en el perfil de daños de una sociedad determinada. Así, además de las clasificaciones tradicionales (enfermedades endémicas, epidémicas y pandémicas, por ejemplo), hoy se habla de enfermedades pretransicionales, transicionales y postransicionales; emergentes y resurgentes, y se ha vuelto común hablar de los perfiles de salud en términos de rezagos o retos epidemiológicos.

En otro terreno, desde hace varias décadas se acepta que el estatuto científico de la salud pública depende, en gran medida, de la cantidad de epidemiología que contenga. Guerra de Macedo, por ejemplo, afirma que las tareas de formar conocimiento nuevo y emplearlo adecuadamente en materia de salud colectiva son específicas de la epidemiología, en especial cuando ésta se concibe "no como un mero instrumento de vigilancia y control de enfermedades, sino en esa dimensión mayor de la inteligencia sanitaria que permite comprender la salud como un todo".[28] La epidemiología, según este punto de vista, no sólo es una parte fundamental de la salud pública, sino muy probablemente su principal fuente de teorías, métodos y técnicas.[29]

**Diseño y evaluación de la respuesta social a los problemas de salud**

La epidemiología también se ha usado como instrumento en la planificación de los servicios sanitarios, mediante la identificación de los problemas prioritarios de salud, las acciones y recursos que son necesarios para atenderlos, y el diseño de programas para aplicar estas acciones y recursos. La evaluación de estos programas —que habitualmente se realiza comparando la frecuencia de enfermedad en el grupo intervenido con la de un grupo testigo y que, por ello, podría denominarse epidemiología experimental—, es un instrumento cada vez más utilizado en el diseño de los planes sanitarios. Así, mediante el uso de métodos y técnicas epidemiológicas se han logrado identificar el impacto real y la calidad con la que se prestan los servicios médicos, así como las formas más eficaces para promover la salud de los que están sanos y las relaciones en-

tre el costo, la efectividad y el beneficio de acciones específicas de salud.

Combinada con otras disciplinas, como la administración, economía, ciencias políticas y ciencias de la conducta, la epidemiología ha permitido estudiar las relaciones entre las necesidades de asistencia y la oferta y demanda de servicios. También con ella se evalúan la certeza de los diversos medios diagnósticos y la efectividad de diferentes terapias sobre el estado de salud de los enfermos. Los estudios sociológicos y antropológicos que hacen uso de técnicas epidemiológicas también son cada vez más frecuentes, y ello ha fortalecido el trabajo y mejorado los resultados de las tres disciplinas.

## ALGUNOS PROBLEMAS EPISTEMOLÓGICOS ACTUALES

La polémica sobre el estatuto científico de la epidemiología fue abierta con la publicación de un controvertido texto elaborado por Carol Buck,[30] en 1975. De acuerdo con esta autora, el hecho de que la epidemiología otorgue tanta importancia a su método se debe a que, en esta disciplina, el experimento juega un papel muy limitado, por lo que los investigadores deben crear escenarios cuasiexperimentales, sirviéndose de los fenómenos tal como ocurren naturalmente. El reconocimiento de esta característica provocó un gran interés en el análisis de los fundamentos lógicos del trabajo epidemiológico, y sus implicaciones epistemológicas se discutieron inmediatamente.[31-33]

En la actualidad, la epidemiología enfrenta varios problemas epistemológicos. De ellos, quizás el más importante es el problema de la causalidad, aspecto sobre el que todavía no existe consenso entre los expertos. El abanico de posturas se extiende desde los que proponen el uso generalizado de los postulados de causalidad (Henle-Koch, Bradford Hill o Evans) hasta los que consideran que la epidemiología debe abandonar el concepto de "causa" y limitarse a dar explicaciones no deterministas de los eventos que investiga. Las críticas al concepto de causa, formuladas por primera vez por David Hume (1740), probablemente implicarían replantear conceptos tan arraigados en la investigación epidemiológica como los de "causa necesaria" y "causa suficiente", por ejemplo. Dado que estas críticas son cada vez más aceptadas en el terreno de las ciencias naturales, es indudable que este tema seguirá siendo uno de los predilectos por la literatura epidemiológica del siglo XXI.

Otro de los problemas filosóficos de la epidemiología contemporánea se refiere a la índole de su objeto de estudio. En este campo, los esfuerzos por determinar la naturaleza de los eventos epidemiológicos también han desembocado en la formación de diversas corrientes, que debaten intensamente si este objeto se alcanza con la suma de lo individual, con el análisis poblacional, o mediante la investigación de lo social. Como resultado, han proliferado los intentos por desentrañar, cada vez con mayor rigor, las interacciones que se establecen entre la clínica, la estadística y las ciencias sociales.[25]

El último de los aspectos centrales en este peculiar debate alude al estatuto científico del saber epidemiológico. Aunque ya nadie acepta la posibilidad —planteada por Louis en el siglo XIX— de que los eventos epidemiológicos puedan comportarse siguiendo leyes similares a las que rigen los fenómenos naturales, los aportes de la epidemiología en el terreno de la generación de teorías, modelos y conceptos han sido numerosos, y su desarrollo presente indica que este proceso no va a detenerse.[34]

## CONCLUSIONES

Es evidente que tanto el objeto como los métodos de estudio de la epidemiología se han modificado radicalmente desde su origen hasta la actualidad. De la descripción del origen sobrenatural de las plagas, la epidemiología ha pasado a explicar la dinámica de la salud poblacional considerada como un proceso social extraordinariamente complejo, compuesto por elementos individuales y colectivos, naturales y culturales, objetivos y subjetivos, caracterizados por poseer una profunda carga histórica y cuyo origen se encuentra tanto en nuestra condición animal como en nuestra naturaleza social.[35] Paulatinamente, la disciplina logró identificar muchos de los elementos que componen este proceso, y de explicar la forma en que se combinan para dar lugar a manifestaciones de aparición regular, cuya comprensión nos permite proponer acciones capaces de modificar efectivamente el curso de su desarrollo.

El avance conceptual de la epidemiología, como ha sucedido desde que nació como ciencia, lejos de detenerse, ha seguido ganando terreno. La teoría de la transición epidemiológica (que desde su nacimiento proporcionó valiosos elementos para interpretar la dinámica de la enfermedad poblacional) ha sido objeto de profundas reformulaciones teóricas.[36] Los conceptos de causa, riesgo, asociación, sesgo, confusión, etc., aunque cada vez son más sólidos, se encuentran en proceso de revisión permanente, lo que hace a la epidemiología una disciplina viva y en constante movimiento.

De acuerdo con Kleinbaum,[37] la nueva epidemiología tiene como propósitos: a) la *descripción* de las condiciones de salud de la población (mediante la caracterización de la ocurrencia de enfermedades, de las frecuencias relativas en el interior de sus subgrupos y de sus tendencias generales); b) la *explicación* de las causas de enfermedad poblacional (determinando los factores que la provocan o influyen en su desarrollo); c) la *predicción* del volumen de enfermedades que ocurrirá, así como su distribución en el interior de los subgrupos de la población, y d) la prolongación de la vida sana mediante el *control* de las enfermedades en la población afectada y la *prevención* de nuevos casos entre la que está en riesgo. Como hemos señalado, también es propósito de la epidemiología *generar* los métodos de abordaje con los cuales puede realizar adecuada y rigurosamente estas tareas.[38] Estos objetivos —que demuestran el avance alcanzado en los dos últimos siglos—indican que, de continuar con la misma tendencia, en las próximas décadas habremos de ver a la disciplina convertida en una ciencia de vastos alcances.



## REFERENCIAS


1. Gill CA. The genesis of epidemics and the natural history of disease. Nueva York: William Wood and Company, 1928:1-39.
2. Cartwright FF, Biddiss M. Disease and history. Nueva York: Thomas Crowell Company, 1972:5-28.
3. Rosen G. A history of public health. Baltimore: The Johns Hopkins University Press, 1958.
4. Sierra J. Obras completas de Justo Sierra. México: UNAM, 1991; vol. 10:33-69.
5. Bucaille M. La Bible, le Coran et la science. París: Editions Seghers, 1987:245-255.
6. La Biblia. Versión de Casiodoro de Reyna (1569). Buenos Aires: Sociedades Bíblicas Unidas, 1960:39-71.
7. Winslow ECA. The conquest of epidemic disea-

se. A chapter in the history of ideas. Madison: Princeton University Press, 1943:117-160.
8. McNeil W. Plagas y pueblos. Madrid: Siglo XXI, 1976:78-146.
9. Sendrail M. Historia cultural de la enfermedad. Madrid: Espasa-Calpe, 1983:57-250.
10. Hipócrates. Hippocratic writings. On airs, waters and places. Chicago: University of Chicago by Encyclopaedia Britannica, 1980:9-19.
11. Kawakita Y, Sakai I, Otzuka M. History of epidemiology. Tokio: EuroAmerica Inc. Publishers, 1993:1-21.
12. Lilienfeld AM, Lilienfeld DE. Fundamentos de epidemiología. México: Addison-Wesley Iberoamericana, 1987:1-38.
13. Ahlbom A, Norell S. Fundamentos de epidemiología. Madrid: Siglo XXI, 1987:VIII-IX.
14. Diccionario etimológico de la lengua castellana. Madrid: Gredos, 1961; vol. 6.
15. Stolley PD, Lasky T. Investigating disease patterns: The science of epidemiology. Nueva York: Scientific American Library, 1995:23-49.
16. Hacking I. La domesticación del azar. Barcelona: Gedisa, 1995:53-112.
17. Foucault M. Historia de la sexualidad. 15ª.ed. México: Siglo XXI, 1987; vol. 1 (La voluntad de saber):168-169.
18. Organización Panamericana de la Salud. El desafío de la epidemiología. Washington: OPS (Publicación Científica núm. 505), 1988:3-17.
19. Hennekens CH H, Buring JE. Epidemiology in medicine. Boston: Little Brown, 1987:73-98.
20. Jenicek M. Epidemiología. Barcelona: Masson,1996:43-78.
21. López-Moreno S, Corcho-Berdugo A, Moreno Altamirano A. Notas históricas sobre el desarrollo de la epidemiología y sus definiciones. Rev Mex Pediatr 1999;66(3):110-114.
22. MacMahon B, Pugh TF. Epidemiology: Principles and methods. Boston: Little Brown, 1970.
23. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. BMJ 1952;2:1271-1286.
24. Rose G. Individuos enfermos y poblaciones enfermas. En: Organización Panamericana de la Salud. El desafío de la epidemiología. Washington: OPS(Publicación Científica núm. 505), 1988:900-909.
25. De-Almeida FN. A clínica e a epidemiologia. Salvador de Bahía: Apce-Abrasco, 1992.
26. Susser M. Choosing a future of epidemiology: From black box to chinese boxes and eco-epidemiology. Am J Public Health 1996;86(5):674-677.
27. Borges G, Medina-Mora ME, López-Moreno S. El papel de la epidemiología en la investigación de los trastornos mentales. Salud Publica Mex 2004;46:451-463.
28. Guerra de Macedo C. Usos y perspectivas de la epidemiología. Washington: Organización Panamericana de la Salud (Publicación Científica núm. 84-47), 1994:6-9.
29. Beaglehole R, Bonita R, Kjellstrom. Epidemiología básica. Washington: Organización Panamericana de la Salud, 1994.
30. Buck C. Popper's philosophy for epidemiologist. Int J Epidemiol 1975;4(3):159-168.
31. Davies, A.M: Comments on "Popper's philosophy for epidemiologists", by Carol Buck. Int J Epidemiol 1975;4(3):169-170.
32. Smith A. Comments on "Popper's philosophy for epidemiologists", by Carol Buck. Int J Epidemiol 1975;4(3):171-172.
33. Jakobsen M. Against popperized epidemiology. Int J Epidemiol 1976;5(1): 9-11.
34. Greenland S. Evolution of epidemiologic ideas. Annotated readings on concepts and methods. 2a. ed. Boston: Epidemiology Resources, 1987.
35. De-Almeida FN. La ciencia tímida. Ensayos de deconstrucción de la epidemiología. Buenos Aires: Lugar Editorial, 2000:83-111.
36. Frenk MJ. La salud de la población. Hacia una nueva salud pública. México: Fondo de Cultura Económica, 1993.
37. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. Nueva York: Van Nostrand Reinhold, 1982.
38. López-Moreno S, Corcho-Berdugo A, López-Cervantes M. La hipótesis de la compresión de la morbilidad: un ejemplo de desarrollo teórico en epidemiología. Salud Publica Mex 1998;40:442-449.

# II Diseño de estudios epidemiológicos

*Mauricio Hernández Ávila*
*Sergio López Moreno*

Los principales objetivos de la investigación epidemiológica son, por un lado, describir la distribución y frecuencia de las condiciones de salud en las poblaciones humanas y, por el otro, contribuir al descubrimiento de los factores ambientales, sociales y biológicos que influyen en estas condiciones. La epidemiología desarrolla las herramientas necesarias para el estudio de todos los eventos relacionados con la salud colectiva, y su propósito es desarrollar conocimiento que pueda ser utilizado para mejorar las condiciones de salud o la manera en que se desarrolla la respuesta social para mantener la salud de la población; por esta razón, se le considera una ciencia básica para la salud pública. La epidemiología, a diferencia de otras ciencias torales de la salud pública, como la nutrición o la salud ambiental, no tiene un campo de conocimientos claramente delimitado. No obstante, sienta las bases para describir la situación de salud de las poblaciones, para diseñar y evaluar programas de salud pública, para evaluar la eficacia de tratamientos a escala poblacional y para investigar y controlar los brotes epidémicos.

Clásicamente, la investigación epidemiológica involucra la recolección y análisis de datos empíricos sobre las poblaciones. Esta información puede generarse ya sea por medio de la experimentación, lo que necesariamente involucra la participación activa e informada de los individuos de la población en estudio, o —lo que es mucho más frecuente— de la observación directa o indirecta de los grupos poblacionales que son objeto de estudio. Para desarrollar un diseño de investigación, ya

**Cuadro I Principales aplicaciones de la epidemiología**

- Descripción de las condiciones del estado de salud de las poblaciones
- Descripción de la historia natural de las enfermedades y otros eventos de salud
- Identificación y caracterización de los factores biológicos, ambientales y sociales que influyen sobre las condiciones de salud
- Identificación de los mecanismos de transmisión y diseminación de las enfermedades
- Identificación y evaluación de factores pronóstico y marcadores tempranos a escala poblacional
- Identificación, descripción y explicación de la frecuencia, distribución, tendencia, vulnerabilidad y formas de satisfacción de las necesidades de salud
- Evaluación de la eficacia, efectividad y confiabilidad de las intervenciones terapéuticas y las medidas diagnósticas a escala poblacional
- Priorización, diseño y evaluación de los programas de salud
- Estudio y control de brotes epidémicos

sea experimental u observacional, es necesario que el epidemiólogo desarrolle estrategias muestrales y de recolección de datos, de medición y de análisis que le permitan estudiar subgrupos poblacionales para, en un segundo momento, hacer extrapolaciones del conocimiento obtenido hacia la población objetivo. La validez de la información derivada de los estudios epidemiológicos depende de lo apropiado de los métodos utilizados. El reconocimiento de la importancia de la metodología para el desarrollo y avance del conocimiento epidemiológico ha propiciado que el desarrollo y el estudio de nuevos métodos de aplicación en el campo se asuman también como un objetivo mismo de la epidemiología. Esto último ha contribuido de manera importante a mejorar la calidad del conocimiento derivado de los estudios epidemiológicos, y a consolidar a la epidemiología como una ciencia básica, necesaria para el avance de la salud pública, de la medicina y de otras ciencias biológicas y sociales.



En este capítulo se revisan las principales estrategias metodológicas utilizadas en la investigación epidemiológica para conformar los grupos poblacionales de estudio. Además, se ofrece un esquema de clasificación ordinal de dichas estrategias en función de la fuerza de la evidencia que aporta cada diseño para establecer relaciones de causa-efecto entre las variables de interés.

A lo largo de los años se han propuesto diferentes esquemas para agrupar y caracterizar los distintos diseños o estrategias de muestreo utilizadas en la investigación epidemiológica. En este trabajo hemos optado por una clasificación multidimensional de acuerdo con: a) el tipo de asignación del factor de estudio considerado como *la exposición*;* b) el número de mediciones que se realiza en cada sujeto de estudio para verificar cambios en la ocurrencia tanto de la exposición como del evento que se considera su *efecto*; c) los criterios que determinan la selección de la población a estudiar; d) la relación temporal entre el inicio del estudio y la medición de la ocurrencia del efecto, y e) la unidad de análisis en la que se miden las variables de interés (cuadro II).[1-5]

La asignación de la variable o factor de estudio que llamamos la exposición es el criterio más importante de clasificación de los estudios epidemiológicos, y se divide en tres tipos: a) experimentales, cuando el investigador controla la exposición utilizando la aleatorización como método de asignación de los sujetos a cada grupo estudiado; b) seudoexperimentales (o de intervención no aleatorizados), cuando el investigador controla la exposición pero no usa procedimientos de aleatorización en la asignación de los sujetos, y c) no experimentales u observacionales, cuando la exposición ocurre sin la intervención del investigador.

De acuerdo con el número de mediciones que se realiza en cada sujeto de estudio para medir la ocurrencia del evento resultado o los cambios observados en la variable de exposición a lo largo del tiempo, los estudios se pueden dividir en: a) longitudinales, cuando se realizan, por lo menos, dos mediciones: una basal, para determinar el estado inicial, y una subsecuente, para determinar si ocurrió o no el evento; y b) transversales, cuando se realiza una sola medición en los sujetos de estudio y, en consecuencia, se evalúan de manera concurrente la exposición y el evento de interés.

En materia de relación causa-efecto, existe una diferencia muy importante entre los estudios longitudinales y transversales, ya que

* En este trabajo utilizaremos el vocablo "exposición" como un término de significado amplio, que puede abarcar desde la exposición a una bacteria o sustancia tóxica hasta la exposición a un suplemento nutricional, una vacuna, un programa de salud o un estilo de vida. En cualquiera de estos casos se presume que la "exposición" es la "causa" de la aparición de un resultado al que puede denominarse, también en forma amplia, como su "efecto" (infección, intoxicación, crecimiento, inmunidad, obesidad, respectivamente).

**Cuadro II Clasificación de los estudios epidemiológicos**

| *Tipo de estudio* | *Asignación de la exposición* | *Número de observaciones por individuo* | *Criterios de selección de la población en estudio* | *Temporalidad del análisis* | *Unidad de análisis* |
|---|---|---|---|---|---|
| Experimental | Controlada (aleatorizada) | Dos o más | Ninguno | Prospectivo | Individuo o grupo |
| Seudoexperimental | Por conveniencia | Dos o más | Ninguno | Prospectivo | Individuo o grupo |
| Cohorte | Fuera del control del investigador | Dos o más | Exposición | Prospectivo o retrospectivo | Individuo |
| Casos y controles | Fuera del control del investigador | Una o más | Efecto | Prospectivo o retrospectivo | Individuo |
| Encuesta | Fuera del control del investigador | Una | Ninguno | Retrospectivo | Individuo |
| Ecológico | Fuera del control del investigador | Dos o más | Ninguno | Retrospectivo | Grupo (o población) |

en los primeros es posible verificar que la exposición antecede al evento de interés, con lo que se cumple el principio temporal de causalidad, según el cual la causa antecede al efecto. En los diseños transversales, en general, es difícil establecer la relación temporal que guardan entre sí las variables, en especial con las exposiciones que varían en el tiempo. Este tipo de estudio, en cambio, proporciona información muy valiosa cuando los factores investigados no varían (como el sexo y la carga genética) o cuando se trata de exposiciones únicas que no cambian con el tiempo, como, por ejemplo, en el caso de la población expuesta a la bomba atómica lanzada en Hiroshima.

La selección de los participantes en el estudio se puede llevar a cabo de acuerdo con la exposición, el evento, o sin considerar ninguna de estas características en los sujetos elegibles para el estudio. La selección con base en estos atributos se utiliza con frecuencia para distinguir entre los diferentes estudios epidemiológicos de tipo observacional. Cuando los sujetos son seleccionados con base en la exposición —es decir, cuando se elige un grupo expuesto y uno no expuesto, en los que posteriormente se determinará la ocurrencia del evento— se considera que se trata de un estudio de cohorte. En contraste, cuando se seleccionan los participantes con base en el evento de estudio —es decir, se elige de manera independiente un grupo de sujetos que tienen el evento de interés (los casos) y un grupo de sujetos que no lo tienen (los controles), y en estos grupos se determina la exposición—, entonces hablamos de un estudio de casos y controles. Finalmente, cuando la selección es indistinta de la ocurrencia de la exposición o del evento —es decir, cuando se seleccionan los sujetos de estudio sin considerar información sobre la exposición o el evento, y la presencia de éstos se determina una vez conformada la población en estudio—, entonces los estudios se denominan transversales, aunque algunos textos también los clasifican como encuestas.[2] La característica principal de esta última estrategia de

muestreo es que la evaluación de la exposición y de la ocurrencia del evento se hacen de manera simultánea. En cuanto a la posibilidad de establecer relaciones de causalidad, se puede decir, de manera general, que los estudios de cohorte tienen mayor peso que los estudios de casos y controles y que los transversales. Sin embargo, es importante mencionar que los estudios de casos y controles, con ciertas características, pueden ser tan informativos (desde el punto de vista de la causalidad) como un estudio de cohorte.

El criterio de temporalidad en la ocurrencia del evento se utiliza para distinguir entre los estudios retrospectivos y prospectivos. El punto de referencia para esta clasificación es la ocurrencia del evento de interés, o variable respuesta. Si al inicio del estudio el evento investigado ya ocurrió y el investigador planea reconstruir su ocurrencia en el pasado, ya sea por medio de registros o entrevistando a los mismos sujetos de estudio, se considera que el estudio es retrospectivo. Si la ocurrencia del evento se registra durante el estudio, es decir, si los sujetos de estudio están libres del evento de interés al iniciar su participación en el estudio, el diseño se considera de tipo prospectivo. En general, puede afirmarse que los estudios prospectivos tienen mayor puntaje en la escala de causalidad, dado que en este tipo de estudios se pueden diseñar instrumentos específicos para el registro y la medición del evento que minimicen los errores de medición y la posibilidad de sesgo. En los estudios retrospectivos, en cambio, la calidad del registro y la medición del evento dependen con frecuencia de la calidad de la información vertida en documentos que no fueron diseñados de manera expresa para observarlo ni para responder a los objetivos de la investigación. Los estudios que incluyen eventos que ocurrieron antes de iniciar la investigación y eventos que aún no han ocurrido al iniciarla son referidos en algunos textos como estudios mixtos o ambispectivos.[3]

Por último, la unidad de análisis se ha utilizado para clasificar los estudios en ecológicos (de conglomerados) e individuales. A diferencia de estos últimos —en los que la unidad de análisis es el individuo y se cuenta, por lo menos, con una medición de cada uno de ellos—, en los estudios ecológicos la unidad de análisis es un grupo (por ejemplo, un país o una región) y se cuenta con el promedio de casos o de exposición para el grupo, y se desconoce la condición de evento o exposición para cada individuo. Debido a que los datos se encuentran agrupados, este tipo de estudios conlleva problemas importantes en su interpretación, ya que no es posible corregir por diferencias en otras variables (posibles variables confusoras) que pudieran explicar los resultados observados. También, dado que la información se encuentra agrupada, no es posible distinguir dentro del conglomerado si la exposición se asocia con la ocurrencia del evento. A esto se debe que este tipo de estudios tenga el peso más bajo en la escala de causalidad.

A continuación describiremos brevemente las características de las principales estrategias epidemiológicas utilizadas para estudiar grupos poblacionales.



## ENSAYOS ALEATORIZADOS

Los ensayos epidemiológicos aleatorizados son estudios experimentales que, cuando se llevan a cabo de manera adecuada, proporcionan el máximo grado de evidencia para confirmar la relación causa-efecto entre la exposición o intervención y el evento en estudio. Se distinguen de los observacionales (no experimentales) porque el investigador tiene control sobre

la asignación de la exposición y porque ésta se lleva a cabo mediante un proceso de aleatorización. Además, dado que se trata de estudios longitudinales y prospectivos completos, y en los que la unidad de análisis es el individuo, es posible prevenir la introducción de sesgos y lograr altos índices de validez. En este tipo de trabajos es posible minimizar la ocurrencia de sesgos mediante una serie de procedimientos que tienen como propósito garantizar la comparabilidad: a) de las intervenciones (considerando que ambos grupos reciben una misma intervención, sólo que el grupo de intervención recibe la que tiene el principio activo que se evalúa), b) de los grupos en estudio, y c) de la información obtenida en la población en estudio.[5]

La comparabilidad de las intervenciones se logra cuando la única diferencia entre los grupos que se comparan es la recepción, por uno de ellos, de la parte activa de la exposición en estudio. El concepto de *comparabilidad* de las intervenciones puede entenderse como una extensión del concepto de *efecto placebo*. Este concepto se deriva del hecho de que el efecto de un medicamento es el resultado de la suma de dos componentes: uno, causado por la sustancia activa del medicamento y, otro, producido por el componente psicológico o social asociado con la idea de recibir un medicamento o con los cambios asociados a la intervención. El propósito de simular una intervención exactamente igual a la que se pretende probar, pero sin la sustancia activa, es precisamente eliminar de la comparación el efecto atribuido al placebo y, de esta manera, estimar únicamente la diferencia atribuible a la sustancia activa o intervención en cuestión.

La comparabilidad de los grupos de estudio se refiere a las características de los participantes y se logra cuando los subgrupos que reciben las diferentes intervenciones son similares en todas y cada una de las características que podrían tener relación con el evento de estudio o con la manera en que actúa la exposición de interés. En términos epidemiológicos, este concepto indica la ausencia de factores de confusión o de modificadores del efecto. En una situación ideal, la comparabilidad de poblaciones podría obtenerse mediante la observación de los mismos sujetos en condiciones experimentales diferentes —con y sin la exposición de interés—. Sin embargo, las condiciones para que esto ocurra son imposibles de observar en la realidad. Como una aproximación para lograr la comparabilidad de poblaciones se ha utilizado la aleatorización. Mediante este proceso se deja al azar la distribución de los sujetos en los diferentes grupos experimentales y se espera que, en promedio, los grupos tengan características comparables, de manera que pueda asegurarse que si se invirtiera la condición de exposición entre los grupos de estudio, los resultados serían equivalentes. Es importante mencionar que la comparabilidad que se obtiene con la asignación aleatoria de los sujetos en los grupos experimentales depende del tamaño muestral, y que la simple aleatorización no garantiza completamente que los individuos se distribuyan homogéneamente en los distintos grupos de intervención. Dado que los individuos se reparten al azar, siempre será necesario verificar el resultado de la aleatorización, debido a que existe la posibilidad de que no funcione adecuadamente. Esto puede llevarse a cabo mediante el cotejo de la distribución de las variables medibles en los individuos que componen los diferentes grupos experimentales: una distribución homogénea entre grupos sería indicativa del éxito de la aleatorización.

Por último, la comparabilidad de la información se logra cuando se utilizan exactamente los mismos métodos de seguimiento y medición de todos los participantes en el estudio. Una de las maneras para lograrlo es no informando a los participantes del estudio respecto del grupo al que pertenecen (al experimental o

al control). También se logra no informando al personal encargado de hacer el seguimiento y mediciones a qué grupo pertenecen los participantes que examinan. Este procedimiento, conocido como enmascaramiento, es simple cuando sólo se lleva a cabo con los pacientes y doble cuando incluye al personal técnico del estudio. Si el evaluador se encuentra enmascarado, es probable que la medición no se vea afectada por esta información. La "ignorancia", en este caso, hace que las mediciones se lleven a cabo de igual forma para todos los individuos. Si por alguna razón se pusiera especial interés en determinar la ocurrencia del evento en cualquiera de los grupos, podría introducirse un sesgo y los resultados podrían ser no comparables. Este concepto abarca tanto la calidad de la información como la proporción de los individuos que pueden no examinarse durante el seguimiento en cada grupo experimental. La pérdida diferencial de participantes entre los grupos que se comparan puede ser el origen de resultados erróneos. Algunos textos refieren este último problema dentro de los sesgos de selección.[2]

Los pasos a seguir para la realización de los estudios experimentales incluyen la definición de la población blanco, que es aquella a la que se pretenden extrapolar los resultados del estudio. Al aplicar los criterios de inclusión en el estudio, se define la población elegible y de allí se seleccionan los participantes; esto último puede llevarse a cabo, ya sea mediante el reclutamiento de voluntarios o mediante la selección de una muestra representativa de la población blanco. Es importante mencionar que siempre que se trabaja con poblaciones humanas se tendrá un grupo autoseleccionado de la población que corresponde a aquellos individuos que otorgan el consentimiento informado* para participar en el proceso experimental. La autoselección de la población tiene gran importancia en cuanto a la validez externa† de los resultados, ya que éstos serán aplicables a la población blanco en la medida en que el grupo en estudio represente adecuadamente a dicha población. Una vez que se han identificado los participantes y que éstos han dado su consentimiento para ser parte del proceso experimental, los individuos se asignan usando un proceso aleatorio a los grupos de estudio. Posteriormente, los grupos se reexaminan en fechas posteriores con el fin de documentar la ocurrencia del evento de interés y los posibles cambios ocurridos en otras covariables (figura 1).

Algunos textos[4] mencionan la existencia de un tipo de estudio adicional entre los estudios experimentales y no experimentales, a los que se les conoce como seudoexperimentales. En estos diseños, el investigador controla la asignación de la exposición; sin embargo, esta asignación no se hace de manera aleatorizada. Este tipo de estudios de intervención incluye los diseños tipo *antes y después* o de *comparación concurrente*.

En los estudios observacionales (no experimentales) la asignación de la exposición ocurre sin la participación del investigador. En este tipo de diseños es común que la exposición ya haya ocurrido al iniciar el estudio, y que ésta se haya dado por algún factor relacionado con las características socioculturales de los participantes. A diferencia de los estudios experimentales, en los que es posible asegurar que la exposición se asigna de manera independiente del evento en estudio, en



* Cuando un paciente acepta ser sometido a una acción médica, firma una carta en la que da su consentimiento. Esto es muy importante en los casos de los procedimientos de investigación, sobre todo, cuando las intervenciones a las que será sometido pueden poner en peligro su salud o su vida. Para otorgar su consentimiento, las personas deben estar bien informadas. Este concepto de consentimiento otorgado se tratará con mayor amplitud en el capítulo 9, relativo a los ensayos clínicos aleatorizados.

† La validez externa es un atributo de los estudios epidemiológicos que permite que sus resultados puedan extrapolarse a la población blanco.

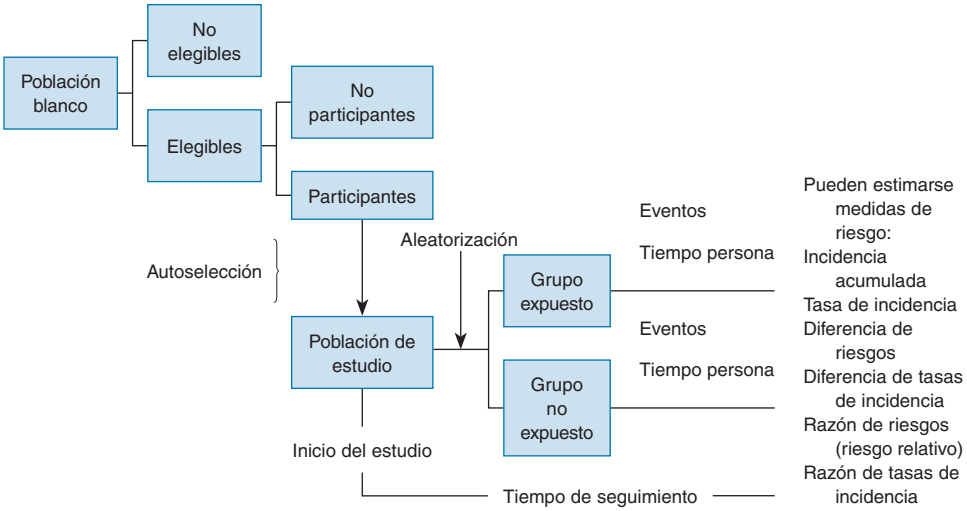



**Figura 1** **Ensayos aleatorizados**

los observacionales esta suposición no es factible. Por esta razón, estos estudios son particularmente susceptibles a los errores que presentan cuando no se cumple el criterio de comparabilidad de poblaciones. También debe evaluarse sistemáticamente la existencia de factores de confusión relacionados con la exposición o el evento en estudio y que, si no se toman en cuenta, pueden introducir sesgos importantes.



## ESTUDIOS DE COHORTE

Entre los estudios observacionales, este tipo de diseño representa lo más cercano al experimental y también tiene un alto valor en la escala de causalidad, ya que es posible verificar la relación causa-efecto correctamente en el tiempo. Sin embargo, dado que se trata de estudios observacionales, tienen la importante limitación de que la asignación de la exposición no es controlada por el investigador ni asignada de manera aleatoria, por lo que no es posible controlar completamente las posibles diferencias entre los grupos expuesto y no expuesto en relación con otros factores asociados con la ocurrencia del evento.

La selección de los participantes con base en la exposición define los estudios de cohorte (figura 2). En este tipo de diseño epidemiológico la población en estudio se define a partir de la exposición y debe estar conformada por individuos en riesgo de desarrollar el evento en estudio. Los sujetos de estudio se seleccionan de la población que tiene la exposición de interés y de grupos poblacionales comparables, pero que no tienen la ex-

posición. Una vez conformada la población en estudio, ésta se sigue en el tiempo y se registra en ella la ocurrencia de la variable respuesta o de otras variables que se consideren de interés.

El diseño de cohorte es especialmente eficiente para estudiar exposiciones raras o poco frecuentes; por ejemplo, las exposiciones ocupacionales que se presentan en poblaciones muy reducidas de trabajadores. En general, cuando se requiere evaluar los riesgos asociados con algún tipo particular de ocupación, se seleccionan grupos ocupacionales y se establece un grupo de comparación (no expuesto) tomado de la población general o, incluso, ubicado en la misma industria o en otra similar, pero no en contacto con la exposición en estudio.

Los estudios de cohorte se utilizan regularmente para estudiar exposiciones que se presentan con una alta frecuencia en la población general. Para este tipo de exposiciones es común seleccionar aleatoriamente grupos representativos de la población que posteriormente se clasifican de acuerdo con la exposición. La cohorte (población en estudio) queda conformada con los participantes que no tienen el evento en estudio y que están en riesgo de desarrollarlo; posteriormente este grupo se sigue en el tiempo con el fin de registrar la ocurrencia del evento. El procedimiento antes descrito se refiere a un estudio prospectivo; sin embargo, los estudios de cohorte también pueden ser retrospectivos. Este segundo tipo de estudios arranca con la definición de los grupos expuesto y no expuesto en algún punto del tiempo



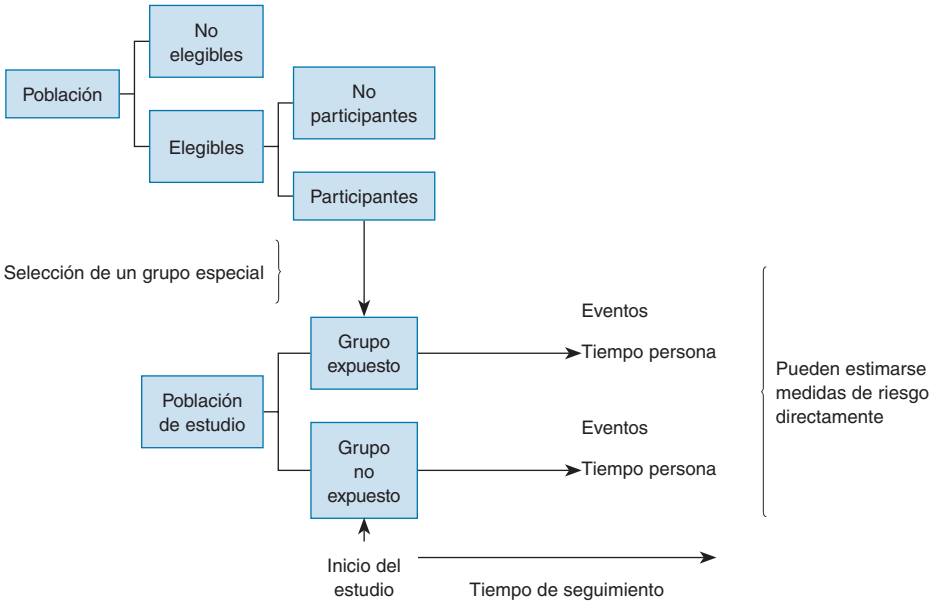



**Figura 2** **Diseños de cohorte prospectiva**

en el pasado, con el fin de reconstruir la experiencia de la cohorte en el tiempo, identificar a los individuos en el tiempo actual o en registros administrativos y evaluar si ya han desarrollado el evento de interés y la fecha en la que lo desarrollaron (figura 3).

En su concepción más simple, un estudio de cohorte consiste en seleccionar un grupo expuesto y otro no expuesto de la población elegible, observarlos durante un tiempo determinado y compararlos en razón de la frecuencia con que ocurre el evento de interés. La validez de la comparación dependerá de que no existan diferencias (aparte de la exposición) entre los grupos expuesto y no expuesto, es decir, de que se cumpla el supuesto de la comparabilidad de poblaciones. Cualquier diferencia en relación con una tercera variable entre ambos grupos y que esté relacionada con la ocurrencia del evento podría distorsionar los resultados sobre la asociación real entre la exposición y el evento (figuras 2 y 3).

Los estudios de cohorte son difíciles de realizar y, además, son costosos. Adicionalmente, este tipo de diseño es poco eficiente para el estudio de enfermedades raras, ya que para registrar un número adecuado de eventos se requiere de un gran número de participantes y tiempos de seguimiento muy prolongados (cuadro III).

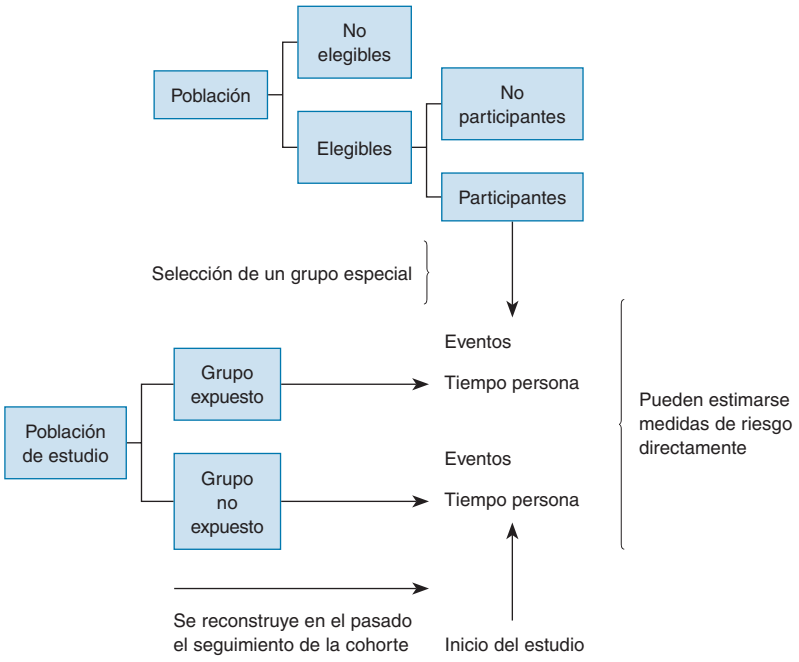



**Figura 3** **Diseños de cohorte retrospectiva**

**Cuadro III Ventajas y desventajas de los estudios de cohorte**

*Ventajas*

- Son los más cercanos a un experimento
- La relación causa-efecto es verificable
- Pueden estimarse medidas de riesgo
- Son eficientes para evaluar exposiciones poco frecuentes
- Pueden estudiarse varios eventos simultáneamente
- Pueden fijarse criterios de calidad en la medición de la exposición y del evento de interés
- Existe un bajo riesgo de sesgos de selección (especialmente en los estudios retrospectivos)

*Desventajas*

- En eventos raros, el costo y el tiempo de seguimiento pueden aumentar considerablemente
- Son difíciles de realizar

## ESTUDIOS DE CASOS Y CONTROLES



Los diseños basados en la selección de los participantes, según tengan o no el evento de interés al momento de ser seleccionados para el estudio, se denominan estudios de casos y controles, o de casos y testigos.[1-5] La principal característica de estos diseños es que la población a investigar se compone de un grupo de individuos con el evento de interés y otro que no lo tiene. Posteriormente, ambos grupos se comparan, buscando identificar la presencia o no de la exposición que se considera asociada al evento.

En los estudios de casos y controles, el investigador fija el número de participantes con el evento de interés (casos) y sin él (controles), que serán incluidos para caracterizar población de referencia que no desarrolló el evento. A diferencia de los estudios de cohorte, en los que se busca igualar la proporción de individuos expuestos y no expuestos en la población de estudio, en este diseño se busca igualar su proporción en cuanto a sujetos con y sin el evento de estudio.

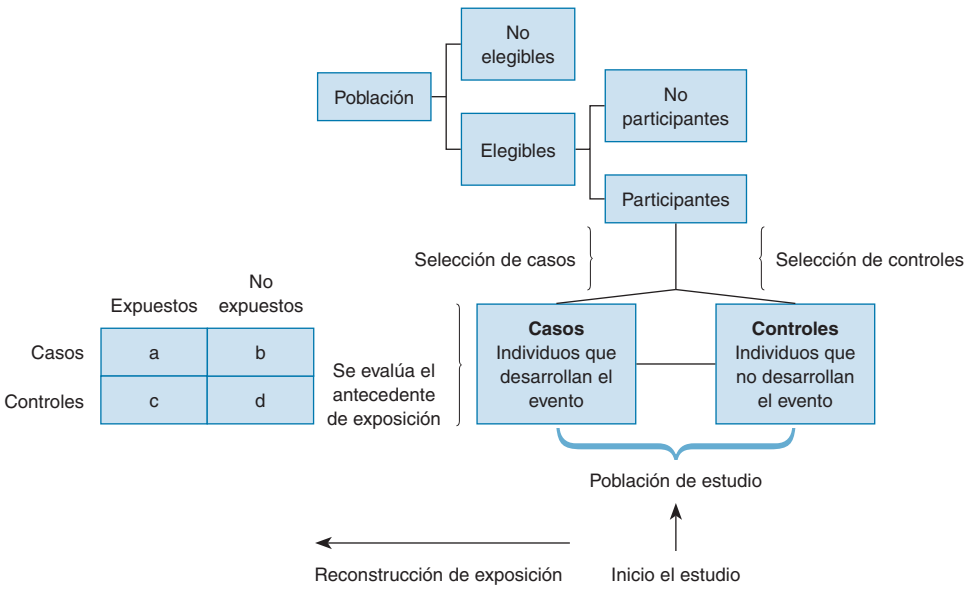
En general, este tipo de estudios se lleva a cabo por medio de registros que permiten identificar fácilmente los sujetos de la población de estudio que desarrollaron el evento: los casos. Los sistemas de registro tradicionalmente utilizados incluyen centros hospitalarios o registros con base poblacional, como los de neoplasias malignas y malformaciones congénitas. El común denominador en este tipo de estudios es la utilización de un sistema que permite concentrar información sobre los casos en tiempos relativamente cortos y, en general, sin la necesidad de invertir los cuantiosos recursos económicos que se requieren para concentrar el mismo número de eventos en un estudio de cohorte.

Los controles se seleccionan de la población que da origen a los casos por medio de un mecanismo independiente del utilizado para seleccionar los casos. La comparación directa de los casos y controles en lo que respecta al antecedente de exposición se utiliza para establecer asociaciones entre la exposición y el evento. Sin embargo, a pesar de que esta comparación frecuentemente se utiliza para establecer de manera automática asociaciones causales entre la exposición y el evento, es muy

importante recalcar que, para que ésta pueda considerarse válida, es indispensable el cumplimiento de ciertas condiciones sobre el origen de los casos y los controles. Entre otras cosas, se requiere que estos últimos tengan la misma base poblacional; que los controles representen de manera adecuada a la población de donde provienen los casos, y que cumplan con la condición de que, si hubieran desarrollado el evento en estudio, habrían participado en el estudio como casos.

Es claro que, a menos que el estudio se desarrolle en el interior de una cohorte bien definida, es difícil verificar el cumplimiento de las condiciones anteriormente mencionadas; empero, el cumplimiento teórico de estas condiciones puede llevarse a cabo si se piensa en este esquema de muestreo como una alternativa para estudiar una cohorte imaginaria, la cual es posible delinear en tiempo y espacio, y hacer el análisis por medio de la selección de una muestra representativa de casos y controles. El supuesto de su origen común se cumple, si tanto casos como controles se originan de la misma cohorte y representan tanto a los casos como a la población en riesgo que no desarrolló el evento.

Los estudios de casos y controles frecuentemente se realizan de manera retrospectiva (figura 4), por lo que no pueden establecer una relación causal perfecta, ya que el evento se evalúa antes que la exposición y, en consecuencia, no siempre es posible verificar que la probable causa antecediera al probable efecto. La naturaleza retrospectiva de los estudios de casos y controles los hace particularmente vulnerables




Sólo puede estimarse de manera directa la Razón de Momios (RM) $RM = (a \cdot d)/(b \cdot c)$

**Figura 4** **Diseños de casos y controles**

**Cuadro IV Ventajas y desventajas de los estudios de casos y controles**

*Ventajas*

- Son eficientes para el estudio de enfermedades raras o periodos de latencia prolongados
- Pueden estudiarse varias exposiciones simultáneamente
- Son menos costosos y requieren menos tiempo que los estudios de cohorte

*Desventajas*

- No es posible estimar directamente medidas de incidencia o prevalencia
- Son susceptibles de sesgos de selección
- Puede presentarse causalidad reversible
- Existen problemas para definir la población de donde provienen los casos
- Existen problemas para medir adecuadamente la exposición

a la introducción de errores en los procesos de selección y recolección de la información. Por esta razón, estos estudios se han colocado en una posición baja en la escala de causalidad. Sin embargo, en ciertas ocasiones, también es posible realizarlos de manera prospectiva y, en ese contexto, tienen mayor peso en la escala de causalidad. Sus principales ventajas y desventajas se describen en el cuadro IV.



## ESTUDIOS TRANSVERSALES

Finalmente, la población en estudio puede seleccionarse de manera aleatoria sin considerar la exposición o el evento como criterios de selección. Este diseño se ha denominado *encuesta transversal* en diferentes textos,[1-4] y se distingue de otros esquemas de muestreo porque en este tipo de estudio se indaga simultáneamente la presencia de la exposición y la ocurrencia del evento una vez conformada la población en estudio (figura 5). El número de eventos y la proporción de individuos expuestos queda determinada por la frecuencia con que éstos ocurren en la población elegible. Esto último contrasta con los estudios de cohorte o de casos y controles, en los que el investigador puede fijar con anterioridad, ya sea la proporción de expuestos (estudio de cohortes) o la del evento en la población en estudio (estudio de casos y controles).

Los estudios transversales se caracterizan porque los sujetos de estudio se evalúan en una sola ocasión, es decir, sólo se hace una medición en cada sujeto; son retrospectivos, y se basan en el estudio de casos prevalentes, que en general representan a los individuos con periodos de mayor sobrevida o duración de la enfermedad. Cualquier factor que se relacione con la duración del evento, la gravedad de la enfermedad y la exposición puede ser una fuente de error en este tipo de trabajos. Este diseño comparte muchas de las limitaciones de los diseños de casos y controles (cuadro V). Por lo anterior, tiene una escala baja en materia de indagación de causalidad, y sus resultados deben interpretarse con mucha cautela. Sin embargo, tales estudios son útiles para la planeación de servicios de salud y para caracterizar el estado de salud de la población en un punto en el tiempo.

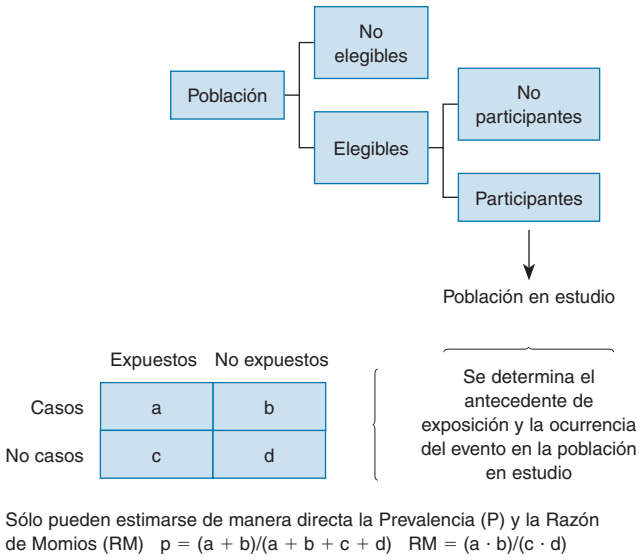



**Figura 5** **Diseño transversal o de encuesta**



## ESTUDIOS ECOLÓGICOS O DE CONGLOMERADOS

En todos los tipos de diseño que hemos mencionado, la unidad de análisis es el participante que compone la población en estudio, y es en éste en el que se mide la exposición y se registra la ocurrencia del evento de interés. Sin embargo, ocasionalmente puede suceder que la unidad de análisis no sea el individuo, y que se utilicen mediciones agregadas en conglomerados o grupos específicos. Dichos conglomerados pueden estar constituidos por grupos poblacionales, comunidades, regiones, estados o países. La característica principal de este tipo de diseños es que se cuenta con información sobre la exposición y el evento para el conglomerado en su totalidad, pero se desconoce la información a escala individual. En este tipo de estudios es común asignar la misma exposición (exposición promedio) a todo el conglomerado, aun cuando se ignore o no se considere la variación individual. Lo mismo sucede con la medición del evento: dado que sólo se cuenta con el número de eventos registrados para el conglomerado, no podemos discernir sobre los eventos que se presentaron en los sujetos expuestos en relación con los no expuestos, por lo que atribuimos la totalidad de eventos de cada conglomerado —sin una verificación real— a la exposición promedio registrada en su interior; posteriormente se comparan el promedio de exposición en cada conglomerado con la frecuencia relativa de los eventos que se presentan en él.

Los estudios ecológicos o de conglomerados permiten estudiar grandes grupos poblacionales en poco tiempo y con un costo relativamente bajo, ya que, en general, se utilizan

**Cuadro V Ventajas y desventajas de los estudios transversales**

*Ventajas*

- Son eficientes para estudiar la prevalencia de enfermedades en la población
- Pueden estudiarse varias exposiciones simultáneamente
- Son baratos y pueden llevarse a cabo en poco tiempo

*Desventajas*

- Existen problemas para identificar y medir adecuadamente la exposición
- Son susceptibles de sesgos de selección
- Son susceptibles de sesgos debidos a casos prevalentes (sobre-representación de los enfermos con mayor sobrevida, con mayor duración de la enfermedad o con manifestaciones clínicas más claras)
- Es posible confundir los factores de riesgo con factores pronóstico y marcadores de enfermedad
- Existe ambigüedad temporal en la relación causa-efecto

**Cuadro VI Ventajas y desventajas de los estudios ecológicos**

*Ventajas*

- Son relativamente sencillos de realizar
- Son baratos y pueden realizarse en poco tiempo
- Permiten estudiar grandes grupos poblacionales o grandes regiones
- Pueden llevarse a cabo por medio de estadísticas vitales y otros registros nacionales

*Desventajas*

- No permiten hacer ajustes por diferencias presentes a escala individual (no es posible identificar cuáles expuestos desarrollaron el evento de interés y cuáles no)
- No se tiene información sobre factores de confusión, por lo que no pueden corregirse los resultados
- No permiten establecer relaciones causa-efecto



estadísticas vitales existentes recolectadas con fines de vigilancia epidemiológica o económica. Sin embargo, dado que ocupan el lugar más bajo en la escala de causalidad, deben considerarse únicamente para sugerir relaciones hipotéticas entre los fenómenos investigados, que tendrán necesariamente que verificarse con estudios más rigurosos. Los principales problemas de este tipo de estudios son que se ignora la variabilidad individual de los integrantes de los conglomerados y que no es posible corregir por diferencias en otras variables que también podrían estar asociadas con la exposición y el evento en estudio (cuadro VI).

## EJEMPLOS DE APLICACIÓN DE LOS PRINCIPALES DISEÑOS

Hemos revisado brevemente los principales diseños de investigación utilizados en los estudios epidemiológicos. Sin duda el ensayo aleatorizado es la estrategia que se reconoce como la más poderosa para establecer relaciones de causa-efecto. Sin embargo, frecuentemente es muy difícil, o imposible, utilizar este tipo de diseño en la investigación epidemiológica, en particular cuando tratamos de evaluar los efectos que tiene la exposición a condiciones que pueden ser comunes en la vida diaria, pero cuyo empleo deliberado en un grupo experimental es éticamente inaceptable. En estas circunstancias, es necesario realizar estudios observacionales en las poblaciones que han estado o están expuestas a las condiciones que creemos

que se asocian a los daños que nos interesa investigar.

Un ejemplo de la situación antes mencionada es la exposición al DDT. Pese a que se consideraría inaceptable la exposición intencional (experimental) de un grupo poblacional a este insecticida, con el único fin de evaluar sus efectos tóxicos, un gran número de personas se expone frecuentemente a esta sustancia, ya sea por actividades ocupacionales o por vivir en regiones donde el DDT se utiliza como un método de control del paludismo. A lo largo de los últimos años, se ha desarrollado un número importante de estudios epidemiológicos observacionales cuyo objetivo principal ha sido el de evaluar si la exposición al DDT de las poblaciones humanas tiene efectos sobre la salud. En particular, se ha evaluado el probable efecto del DDT sobre el desarrollo de cáncer. Mencionaremos algunos de ellos con el fin de ilustrar la utilización de los diferentes diseños epidemiológicos descritos anteriormente en un problema de actualidad en el campo de la salud pública.

Cocco y colaboradores[6] realizaron un estudio de cohorte retrospectivo con el fin de estudiar el riesgo de desarrollar cáncer por exposición laboral a DDT. Para este estudio se definió como población expuesta a los sujetos que participaron en el programa de control del paludismo en Sardenia, Italia. Los investigadores obtuvieron información sobre el estado de salud y la causa de muerte de los trabajadores que participaron en las actividades de control, grupo en el que se registraron 1 043 muertes. La distribución y frecuencia de las causas de muerte observada se compararon con las que se registraron en la población italiana del mismo sexo y edad, considerada como el grupo no expuesto.

Hunter y colaboradores[7] estudiaron la asociación entre los niveles séricos de DDE (indicador de exposición a DDT) y la incidencia de cáncer mamario. Los investigadores realizaron un estudio de casos y controles prospectivo en una cohorte bien definida de enfermeras estadounidenses, iniciado en 1986. En la población de estudio incluyeron 240 casos de cáncer mamario que se registraron en una cohorte conformada con enfermeras en el año 1986. Los autores registraron los casos de cáncer de mama que ocurrieron entre 1990 y 1992. La concentración sérica de DDE se comparó con la observada en un grupo de participantes de la misma cohorte y del mismo grupo de edad, pero que no desarrollaron cáncer mamario. La concentración de DDE se determinó en muestras de suero que habían sido recolectadas en años previos al diagnóstico de cáncer mamario.

Romieu y colaboradores[8] realizaron un estudio de casos y controles para estudiar la asociación entre la exposición ambiental a DDT y el riesgo de desarrollar cáncer mamario en mujeres residentes de la Ciudad de México. En este trabajo, los investigadores estudiaron una serie de 120 casos incidentes de cáncer mamario reclutados en unidades médicas del Instituto Mexicano del Seguro Social, el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado y la Secretaría de Salud, y compararon de manera concurrente los niveles de exposición a DDT con los observados en una serie de 123 controles que se reclutaron mediante un muestreo aleatorio de mujeres de la Ciudad de México.

Wang y colaboradores[9] realizaron un estudio multinacional de conglomerados para estudiar la asociación entre exposición al DDT y cáncer. En este estudio, los autores recuperaron muestras de cerumen de 3 800 personas entre 35 y 54 años de edad de 35 países, y determinaron las concentraciones promedio de exposición a DDT en las muestras de cerumen. La concentración promedio de DDT fue utilizada para caracterizar la exposición promedio de cada país al DDT y posteriormente correlacionaron esta exposición con la tasa

de mortalidad por cáncer de los países incluidos en el estudio.

Finalmente, Ayotte y colaboradores[10] estudiaron la asociación entre los niveles plasmáticos de DDT y algunos parámetros de capacidad reproductiva mediante un estudio transversal. Para fines de este estudio, los investigadores reclutaron voluntarios de una zona palúdica de México y determinaron de manera transversal la asociación entre los niveles séricos de DDT y el número de espermatozoides, así como de otros parámetros funcionales. En este estudio también se evaluó transversalmente la asociación entre la concentración plasmática de DDT y la de diferentes proteínas transportadoras de hormonas sexuales masculinas.

Los diseños epidemiológicos descritos en este trabajo representan los tipos más utilizados. La aplicación de cada método a la resolución de diferentes problemas de salud requiere de creatividad y conocimiento de los alcances y limitaciones de cada uno de ellos, así como de los métodos de análisis que se han desarrollado de manera específica para cada aplicación. En los capítulos subsecuentes se presentará con mayor detalle cada uno de estos diseños.

## REFERENCIAS


1. Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill: Epidemiology Resources Inc., 1991.
2. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. Nueva York: Oxford University Press, 1986.
3. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Belmont: Lifetime Learning Publications, 1982.
4. Rothman KJ, Greenland S. Modern epidemiology. 2a. ed. East Washington Square: Lippincott-Raven Publishers, 1998.
5. Miettinnen OS. Theoretical epidemiology. Principles of occurrence research medicine. Nueva York: A Wiley Medical Publication, 1985.
6. Cocco P, Blair A, Congia P, Saba G, Ecca AR, Palmas C. Long-term health effects of the occupational exposure to DDT. A preliminary report. Ann NY Acad Sci 1997;837:246-256.
7. Hunter DJ. Hankinson SE. Laden F. Colditz GA. Manson JE. Willett WC *et al*. Plasmaorganochlorine levels and the risk of breast cancer. N Engl J Med 1997;337(18):1253-1258.
8. Romieu I, Hernández-Ávila M, Lazcano-Ponce E, Weber JP, Dewali E. Breast cancer, lactation history and serum organochlorines. Am J Epidemiol 2000;152(4):363-370.
9. Wang XQ. Gao PY. Lin YZ. Chen CM. Studies on hexachlorocyclohexane and DDT contents in human cerumen and their relationships to cancer mortality. Biomed Environ Sci 1988;1(2):138-151.
10. Ayotte P, Giroux S, Dewailly E, Hernández-Ávila M, Farías P, Danis R, Villanueva-Díaz CA. DDT spraying for malaria control and reproductive function in mexican men. Epidemiology. 2001;12(3):366-367.

# III Principales medidas

*Alejandra Moreno Altamirano*
*Sergio López Moreno*
*Mauricio Hernández Ávila*

Todo proceso de investigación inicia con la identificación de un problema científico y el planteamiento de una hipótesis que, al ponerse a prueba, lo responda tentativamente. Para contrastar la hipótesis se requiere descomponerla en un conjunto suficientemente pequeño de variables, susceptibles de ser evaluadas empíricamente. Si los procedimientos empíricos no refutan la hipótesis planteada, ésta se acepta como probablemente verdadera. En forma muy resumida, éste es el camino que el científico sigue más frecuentemente al realizar su trabajo. Dado que en la mayoría de los casos es necesario medir las variables durante la contrastación empírica de la hipótesis, la medición resulta un procedimiento muy frecuente en la práctica científica.

En epidemiología, el proceso de investigación es similar al utilizado en el resto de las ciencias. Cuando se investiga la salud de la población también se proponen una o varias explicaciones hipotéticas que posteriormente son sometidas a contrastación empírica. En este proceso, los conceptos de *medición* y de *variable* resultan fundamentales.



## CONCEPTO DE MEDICIÓN

La medición consiste en asignar un número o calificación a una propiedad específica de un individuo, una población o un proceso, usando ciertas reglas. La principal característica de la medición es que siempre consta de una fase de abstracción y una de operación. La primera es tan indispensable como la segunda, pues en términos estrictos nunca es posible medir un proceso o una cosa en forma completa, sino sólo cierta característica suya, abstrayéndola de otras propiedades y bajo la influencia de las limitaciones del método utilizado. Por ejemplo, uno no mide el desarrollo de un niño o su estado nutricional, sino que obtiene información sobre su estatura o su peso. La fase de operación radica en comparar el atributo medido contra una escala considerada estándar, con el fin de evaluar sus cambios en el tiempo o en el espacio. El proceso de medir, en consecuencia, implica el paso de una entidad teórica a una escala conceptual y, posteriormente, a una escala operativa.

En general, los pasos que se siguen durante la medición son los siguientes: a) se delimita la parte del evento que se medirá, b) se selecciona la escala con la que se va a medir, c) se compara el atributo medido con la escala, y d) finalmente, se emite un juicio de valor acerca de los resultados de la comparación. Para medir el crecimiento de un niño, por ejemplo, primero se seleccionan las variables a medir (edad, peso, talla); luego se seleccionan las unidades

de medición (meses cumplidos o días, centímetros o milímetros, kilos o gramos); enseguida se comparan los atributos del individuo con las escalas seleccionadas, y se les otorga un valor numérico (un mes de edad, 60 cm de talla, 4 500 gr de peso) y, por último, se emite un juicio de valor, que sintetiza la comparación entre las magnitudes encontradas y los criterios de salud aceptados como válidos en ese momento. Como resultado, el infante se califica como bien nutrido, desnutrido o con sobrepeso. También es posible relacionar los valores encontrados para la variable dependiente con los documentados para la exposición y, de esta manera, contrastar la hipótesis planteada.

Como puede notarse, la medición es un proceso instrumental sólo en apariencia, ya que la selección de la parte que se medirá de la escala de medición y de los criterios de salud que se usarán como elementos de juicio deben ser resultado de un proceso de decisión teórica. La medición nos permite alcanzar un alto grado de objetividad al usar los instrumentos, escalas y criterios aceptados como válidos y que pueden ser reproducidos y constatados por otros investigadores.

## CONCEPTO DE VARIABLE

La función de las variables consiste en proporcionar información asequible para descomponer la hipótesis planteada en sus elementos más simples. Las variables pueden definirse como aquellos atributos o características de los eventos, de las personas o de los grupos de estudio que cambian de una persona a otra o de un tiempo a otro en la misma persona y que, por lo tanto, pueden tomar diversos valores. Para su estudio es necesario medirlas en el objeto investigado, y es en el marco del problema y de las hipótesis planteadas donde adquieren el carácter de variables.

De acuerdo con la relación que guardan unas con otras en el contexto de la hipótesis que se evalúa, las variables se clasifican en independientes y dependientes. Cuando se supone que una variable produce un cambio en otra, la primera se considera independiente y la segunda, dependiente o de respuesta. En los estudios epidemiológicos, la enfermedad o evento es por lo general la variable dependiente o de respuesta; los factores que determinan su aparición, magnitud y distribución son las independientes. En la literatura epidemiológica también se utiliza con frecuencia el término exposición para identificar las variables independientes. No obstante, el concepto de dependencia e independencia es contextual, es decir, obedece al modelo teórico planteado para la hipótesis en cuestión. Una vez que se han identificado las variables conceptuales, el investigador debe definirlas de manera operativa, especificando el método y la escala de medición.

El uso de variables permite la elaboración de modelos descriptivos, explicativos y predictivos sobre la dinámica de la salud poblacional. En los modelos más sencillos (por ejemplo, en los modelos en los que se considera una sola exposición con dos niveles: presente/ausente y un solo daño, medido en dos niveles: presente/ausente) las variables generalmente se expresan en cuadros simples de dos categorías mutuamente excluyentes (llamadas dicotómicas), representadas por la ausencia y la presencia de la exposición, y la ausencia y la presencia del daño. Al combinar ambas categorías se forma un cuadro con dos filas y dos columnas, conocido como tabla tetracórica, de contingencia o tabla de $2 \times 2$. Cuando, en cambio, existen más de dos categorías de ex-

posición, o varias formas de clasificar el evento, esta relación se expresa en cuadros de varias columnas y varias celdas. En este capítulo se analizará la elaboración de medidas epidemiológicas basadas en categorías dicotómicas y el uso de tablas de 2 × 2.

## PRINCIPALES ESCALAS

Las escalas se clasifican en cualitativas (nominal y ordinal) y cuantitativas (de intervalo y de razón). Un requisito indispensable en todas las escalas es que las categorías deben ser exhaustivas y mutuamente excluyentes. En otras palabras, debe existir una categoría para cada caso que se presente y cada caso debe poder colocarse en una sola categoría.

### Escala nominal

La medición de carácter nominal consiste simplemente en clasificar las observaciones en categorías diferentes con base en la presencia o ausencia de cierta cualidad. De acuerdo con el número de categorías resultantes, las variables se clasifican en dicotómicas (dos categorías) o politómicas (más de dos categorías). En las escalas nominales no es posible establecer un orden de grado como mejor o peor, superior o inferior, o mayor o menor. La asignación de códigos numéricos a las categorías se hace con el único fin de diferenciar unas de otras y no tienen interpretación en lo que se refiere al orden o magnitud del atributo. Como ejemplos de este tipo de medición en la investigación epidemiológica se pueden mencionar el género (masculino, femenino), el estado civil (soltero, casado, viudo, divorciado), la exposición o no a un factor, y el lugar de nacimiento, entre otras.

### Escala ordinal

En contraste con las escalas nominales, en este tipo de medición las observaciones se clasifican y ordenan por categorías según el grado en que los objetos o eventos poseen una determinada característica. Por ejemplo, las personas pueden clasificarse de acuerdo con el grado de una enfermedad en: sanos, con enfermedad leve, con enfermedad moderada o con enfermedad severa. Si se utilizan números en este tipo de escalas, su única significación consiste en indicar la posición de las distintas categorías de la serie y no la magnitud de la diferencia entre las categorías. Para la variable antes mencionada, por ejemplo, sabemos que existe una diferencia de grado entre leve y severo; pero no es posible establecer con exactitud numérica la magnitud de la diferencia en la gravedad con la que se manifiesta la enfermedad de una u otra persona. Esta característica hace que ciertas comparaciones numéricas también encuentren limitaciones con el uso de este tipo de escalas.

### Escala de intervalo

La escala de intervalo es de tipo cuantitativo. Además de ordenar las observaciones por categorías del atributo, mide la magnitud de la distancia relativa entre las categorías; sin embargo, no proporciona información sobre la magnitud absoluta del atributo medido. Por ejemplo, se obtiene una escala de intervalo para la altura de las personas de un grupo cuando, en lugar de medirlas directamente, se mide la altura de cada persona respecto a la altura promedio. En este caso, el valor cero es arbitrario y los valores asignados a la altura no expresan su magnitud absoluta. Esta es la característica distintiva de las escalas de intervalo en comparación con las de razón.

El ejemplo más conocido de las escalas de intervalo es la utilizada para medir la tempe-

ratura. Por convención, el grado cero corresponde al punto de congelación del agua y, por lo tanto, la razón entre dos objetos con temperaturas de, por ejemplo, 40 y 80 grados, no indica que uno de ellos sea realmente dos veces más caliente (o más frío) que el otro. En ciencias de la salud, un ejemplo de este tipo de escalas es la utilizada para medir el coeficiente intelectual.

### Escalas de razón

La escala de razón tiene la cualidad de que el cero sí indica la ausencia del atributo y, por lo tanto, la razón entre dos números de la escala es igual a la relación real existente entre las características de los objetos medidos. En otras palabras, cuando decimos que un objeto pesa 8 kilogramos, también estamos diciendo que pesa el doble que otro cuyo peso es de 4 kilogramos. Muchas características biofísicas y químicas que pueden medirse en las unidades convencionalmente aceptadas (metros, gramos, micras, molécula/kilogramo, miligramo/decilitro, etc.) son ejemplos de mediciones que corresponden a este tipo de escala. En materia de investigación de salud, el ingreso económico y la concentración de plomo en la sangre son ejemplos de este tipo de escalas.

## CÁLCULO DE PROPORCIONES, TASAS Y RAZONES



Un método frecuentemente utilizado en los estudios epidemiológicos para la contrastación de hipótesis es el método probabilístico. Este método postula que las relaciones causales hipotetizadas entre las variables, se pueden evaluar en términos probabilísticos. Es decir, se trata de establecer si la mayor o menor probabilidad de que un evento ocurra se debe a los factores que se sospecha intervienen en su génesis y no debido a la variación biológica natural que se observa en los diferentes fenómenos, es decir al azar. Para cumplir con este objetivo, la investigación epidemiológica utiliza diferentes modelos probabilísticos que se ajustan principalmente a tres tipos de medidas: a) de frecuencia; b) de asociación o efecto, y c) de impacto potencial. La construcción de estas medidas se realiza por medio de operaciones aritméticas que corresponden a razones, proporciones y tasas. Antes de abordar las medidas utilizadas en los estudios epidemiológicos revisaremos brevemente las características de estas tres últimas operaciones.

### Proporciones

Las proporciones son medidas que expresan la frecuencia con la que ocurre un evento en relación con la población total. Esta medida se calcula dividiendo el número de eventos ocurridos entre la población en la que ocurrieron. Como cada elemento de la población puede contribuir únicamente con un evento, es lógico que al ser el numerador (el volumen de eventos) una parte del denominador (población en la que se presentaron los eventos), el primero nunca será más grande que el segundo. Esta es la razón por la que el resultado no puede ser mayor que la unidad y oscila siempre entre cero y uno. Por ejemplo, si en un año se presentan tres muertes en una población compuesta por cien personas, la proporción anual de muertes en esa población será:

$$P = \frac{3 \text{ muertes}}{100 \text{ personas}} = 0.03$$

A menudo, las proporciones se expresan en forma de porcentaje, y en tal caso los resultados oscilan entre cero y cien. En el ejemplo

anterior, la proporción anual de muertes en la población sería de 3 por 100, o de 3%. Nótese, asimismo, que el denominador no incluye el tiempo. Las proporciones expresan únicamente la relación que existe entre el número de veces en las que se presenta un evento y el número total de ocasiones en las que pudo presentarse.

**Tasas**

Las tasas expresan la dinámica de un suceso en una población a lo largo del tiempo. Se pueden definir como la magnitud del cambio de una variable (enfermedad o muerte) por unidad de cambio de otra (usualmente el tiempo) en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso. En las tasas, el numerador expresa el número de eventos acaecidos durante un periodo en un número determinado de sujetos observados.

A diferencia de una proporción, el denominador de una tasa no expresa el número de sujetos en observación, sino el tiempo durante el cual tales sujetos estuvieron en riesgo de sufrir el evento. La unidad de medida empleada se conoce como tiempo-persona de seguimiento u observación. Por ejemplo, la observación de cien individuos en riesgo de padecer el evento durante un año corresponde a cien años-persona de seguimiento; de manera similar, diez sujetos observados durante diez años corresponden a cien años-persona de seguimiento.

Dado que el periodo entre el inicio de la observación y el momento en que aparece un evento puede variar de un individuo a otro, el denominador de la tasa se estima a partir de la suma de los periodos de todos los individuos. Las unidades de tiempo pueden ser horas, días, meses o años, dependiendo de la naturaleza del evento que se estudia.

El cálculo de tasas se realiza dividiendo el total de eventos ocurridos en un periodo dado en una población entre el tiempo-persona total (es decir, la suma de los periodos individuales libres de la enfermedad) en el que los sujetos estuvieron en riesgo de presentar el evento. Las tasas se expresan multiplicando el resultado obtenido por una potencia de 10, con el fin de permitir rápidamente su comparación con otras tasas.

$$\text{Tasa} = \frac{\text{número de eventos ocurridos en una población en un periodo } t}{\text{sumatoria de los periodos durante los cuales los sujetos de la población libres del evento estuvieron expuestos al riesgo de presentarlo en el mismo periodo}} \times \text{una potencia de 10}$$

**Razones**

Las razones pueden definirse como magnitudes que expresan la relación aritmética existente entre dos eventos en una misma población, o un solo evento en dos poblaciones. En el primer caso, un ejemplo es la razón de residencia hombre:mujer en una misma población. Si en una localidad residen 5 000 hombres y 4 000 mujeres se dice que, en ese lugar, la razón de residencia hombre:mujer es de 1:0.8 (se lee uno a 0.8), lo que significa que por cada hombre residen ahí 0.8 mujeres. Esta cantidad se obtiene como sigue:

$$\text{Razón hombre:mujer} = \frac{4\,000}{5\,000} = 0.8$$

En este caso, también podría decirse que la razón hombre:mujer es de 10:8, pues esta expresión aritmética es igual a la primera (1:0.8).

En el segundo ejemplo se encuentran casos como la razón de tasas de mortalidad por causa específica (por ejemplo, por diarreas) en dos comunidades. En este caso, la razón expresaría la relación cuantitativa que existe entre la tasa de mortalidad secundaria a diarreas registrada en la primera ciudad, y la tasa de mortalidad secundaria a diarreas registrada en la segunda. La razón obtenida expresa la magni-

tud relativa con la que se presenta este evento en cada población. Si la tasa de mortalidad por diarreas en la primera ciudad es de 50 por 1 000 y en la segunda de 25 por 1 000, la razón de tasas entre ambas ciudades sería:

$$\text{RTM} = \frac{\text{tasa de mortalidad en la ciudad B}}{\text{tasa de mortalidad en la ciudad A}} = \frac{50 \times 1\,000}{25 \times 1\,000} = 2.0$$

Donde RTM es la razón de tasas de mortalidad (en este caso, por diarreas) entre las ciudades A y B. El resultado se expresa como una razón de 1:2, lo que significa que por cada caso en la ciudad A hay dos en la ciudad B.

## MEDIDAS DE FRECUENCIA

El paso inicial en gran parte de las investigaciones epidemiológicas es medir la frecuencia de los eventos de salud con el fin de hacer comparaciones entre distintas poblaciones o en la misma población a través del tiempo. No obstante, dado que el número absoluto de eventos depende en gran medida del tamaño de la población en la que se investiga, estas comparaciones no pueden realizarse con cifras de frecuencia absoluta (o número absoluto de eventos).

Por ejemplo, si en dos diferentes poblaciones se presentan 100 y 200 casos de cáncer cérvico-uterino, respectivamente, podría pensarse que en el segundo grupo la magnitud del problema es del doble que en el primero. Sin embargo, esta interpretación sería incorrecta si el segundo grupo tuviera el doble de tamaño que el primero, ya que la diferencia en el número de casos, simplemente, podría deberse al mayor tamaño de la segunda población y no a la presencia de una tercera variable que pudiera explicar estas diferencias. En general los cambios en la frecuencia absoluta de los eventos no son informativos.

En consecuencia, para comparar adecuadamente la frecuencia de los eventos de salud es necesario construir una medida que considere el tamaño de la población en la que se realiza la medición. Este tipo de medidas, denominadas medidas de frecuencia relativa, se obtiene, en general, relacionando el número de casos (numerador) con el número total de individuos que componen la población (denominador). El cálculo correcto de estas medidas requiere que se especifique claramente cómo se construyen el numerador y el denominador. Así, la tasa de mortalidad por cáncer cérvico-uterino se construye dividiendo el número de muertes por esta causa en mujeres mayores de 25 años o más por 100 000 mujeres de este grupo de edad. Es evidente, por ejemplo, que los varones no deben ser incluidos en el denominador durante el cálculo de la frecuencia relativa de esta enfermedad.

La población que es susceptible a una enfermedad se denomina población en riesgo. Así, por ejemplo, los accidentes laborales sólo pueden ocurrir en las personas que trabajan, por lo que la población en riesgo es la población trabajadora. Si, en cambio, queremos investigar el efecto de un contaminante generado por una fábrica, podríamos ampliar el denominador a toda la población expuesta al mismo, sea o no trabajadora.

Las medidas de frecuencia más usadas en epidemiología se refieren a la medición de la mortalidad o la morbilidad (especialmente la

incidencia, la prevalencia y la duración de la enfermedad). La mortalidad es útil para estudiar enfermedades que provocan la muerte. Empero, cuando la letalidad es baja y, en consecuencia, la frecuencia con la que se presenta una enfermedad no puede analizarse adecuadamente con los datos de mortalidad, las mediciones de la morbilidad (incidencia, prevalencia y duración de la enfermedad) se convierten en las medidas epidemiológicas de mayor importancia.

En ocasiones, la morbilidad también puede servir para explicar las tendencias de la mortalidad, ya que los cambios en las tasas de mortalidad están íntimamente relacionados con cambios en la duración de la enfermedad o con la prevalencia de la misma. Por ejemplo, la disminución en la mortalidad infantil explica los aumentos aparentes que ocurren posteriormente en el volumen de enfermedades en otras edades; la prevalencia de infecciones por VIH también puede incrementarse debido a una mejoría en la sobrevida atribuida a la introducción de nuevos medicamentos. Por ambas razones, el análisis de las condiciones de salud de las poblaciones generalmente se basa en los cambios observados en las medidas de mortalidad y morbilidad.

Las principales fuentes de información de morbilidad son los datos hospitalarios, los registros poblacionales de enfermedades y las encuestas; sin embargo, debido a que en muchas ocasiones estos registros tienen limitaciones, los estudios epidemiológicos frecuentemente utilizan información obtenida mediante métodos especialmente diseñados para ello. A continuación se presenta un resumen de los elementos más importantes de las medidas de mortalidad y morbilidad.

## MEDIDAS DE MORTALIDAD



El concepto de mortalidad expresa la magnitud con la que se presenta la muerte en una población en un lapso de tiempo determinado. A diferencia de los conceptos de muerte y defunción, que reflejan la pérdida de la vida biológica individual, la mortalidad es una categoría de naturaleza estrictamente poblacional. En consecuencia, la mortalidad expresa la dinámica de las muertes acaecidas en las poblaciones a través del tiempo y el espacio, y sólo permite comparaciones en este nivel de análisis. La mortalidad puede estimarse para todos o algunos grupos de edad, para uno o ambos sexos, y para una, varias o todas las enfermedades. La mortalidad se clasifica de la siguiente manera: a) general y b) específica.

### Mortalidad general

La mortalidad general es el volumen de muertes ocurridas por todas las causas de enfermedad, en todos los grupos de edad y para ambos sexos. La mortalidad general, que comúnmente se expresa en forma de tasa, puede ser cruda o ajustada, de acuerdo con el tratamiento estadístico que reciba. Esto último será necesario siempre que se comparen distintas poblaciones.

La mortalidad cruda expresa la relación que existe entre el volumen de muertes ocurridas en un periodo dado y el tamaño de la población en la que se presentaron; la mortalidad ajustada (o estandarizada) expresa esta relación, pero considera las posibles diferencias en términos de una tercera variable, como podría ser: la estructura por edad, género, o cualquier otra variable. La estandarización permite hacer comparaciones válidas entre diferentes poblaciones. En este caso, las tasas se reportan como tasas ajustadas o estandarizadas. La tasa cruda de mortalidad se calcula de acuerdo con la siguiente fórmula:

$$\text{Tasa mortalidad general} = \frac{\text{número de muertes en el periodo } t}{\text{población total promedio en el mismo periodo}} = (\times 10n)$$

**Mortalidad por edad o causa específica**

Cuando existen razones para suponer que la mortalidad puede variar entre los distintos subgrupos de la población, ésta se divide para su estudio. Cada una de las medidas obtenidas de esta manera adopta su nombre según la fracción poblacional que se reporte. Por ejemplo, si las tasas de mortalidad se calculan para los diferentes grupos de edad, serán denominadas tasas de mortalidad por edad. De la misma manera pueden calcularse la mortalidad por sexo o por causa específica, etcétera.

En algunos casos pueden calcularse para subgrupos específicos de edad y sexo (por ejemplo, mortalidad femenina en edad reproductiva, mortalidad materna). Las tasas de mortalidad específica por edad y sexo se calculan de la siguiente forma:

$$\text{TME} = \frac{\text{total de muertes en un grupo de edad y sexo específicos de la población durante un periodo dado}}{\text{población total estimada del mismo grupo de edad y sexo en el mismo periodo en el mismo periodo}} = (\times 10n)$$

Donde TME es la tasa de mortalidad específica para esa edad y sexo.

*Tasa de letalidad.* Es una medida de la gravedad de una enfermedad, considerada desde el punto de vista poblacional, y se define como la proporción de casos de una enfermedad que resultan mortales respecto del total de casos que contrajeron la enfermedad en un periodo especificado. La medida indica la importancia de la enfermedad en términos de su capacidad para producir la muerte y se calcula de la manera siguiente:

$$\text{Letalidad (\%)} = \frac{\text{número de muertes por una enfermedad en un periodo}}{\text{número de casos diagnosticados de la misma enfermedad en el mismo periodo}} \times 100$$

La letalidad, en sentido estricto, es una proporción, ya que expresa el número de defunciones entre el total de casos diagnosticados con la enfermedad. No obstante, generalmente se expresa como tasa de letalidad y se reporta como el porcentaje de muertes de una causa específica respecto del total de enfermos de esa causa durante un periodo determinado.



## MEDIDAS DE MORBILIDAD

La enfermedad puede medirse en términos de prevalencia o de incidencia. La prevalencia se refiere al número de individuos que, en relación con la población total, padecen una enfermedad determinada en un momento específico. Debido a que un individuo sólo puede encontrarse sano o enfermo en relación con cualquier enfermedad, la prevalencia representa la probabilidad de que un individuo sea un caso de dicha enfermedad en un momento específico.

La incidencia, por su parte, expresa el volumen de casos nuevos que aparecen en un periodo determinado entre la población en riesgo; también puede manifestar la velocidad con que aparecen tales casos. Es decir, las diferentes medidas de incidencia pueden expresar tanto la probabilidad de enfermar por parte de una población en riesgo como la velocidad con la que los individuos desarrollarán una enfermedad determinada durante cierto periodo.

## Prevalencia

La prevalencia es una proporción que indica la frecuencia de un evento. En general, se define como la proporción de la población que padece la enfermedad en estudio en un momento dado, y se denomina únicamente como prevalencia ($p$). Como todas las proporciones, no tiene dimensiones y nunca puede tomar valores menores de 0 o mayores de 1. A menudo se expresa como casos por mil o por cien habitantes. Si los datos se han recogido en un momento o punto temporal dado, $p$ es llamada prevalencia puntual.

*Prevalencia puntual.* Es la probabilidad de un individuo de una población de tener el evento en estudio en el momento $t$, y se calcula de la siguiente manera:

$$p = \frac{\text{número total de casos existentes en el momento } t}{\text{total de la población estudiada en el momento } t} (\times 10\text{n})$$

La prevalencia de una enfermedad aumenta como consecuencia de una mayor duración de la enfermedad, el aumento de casos nuevos, la inmigración de casos (o de susceptibles), la emigración de sanos y la mejoría de las posibilidades diagnósticas o terapéuticas. La prevalencia de una enfermedad disminuye cuando es menor la duración de la enfermedad, existe una elevada tasa de letalidad, disminuyen los casos nuevos, hay inmigración de personas sanas, emigración de casos y aumento de la tasa de curación. En resumen, la prevalencia de una enfermedad depende de la incidencia y de la duración de la enfermedad. Por esta razón, aunque con muchas limitaciones, la prevalencia puntual puede ser un buen indicador de la incidencia de la enfermedad.

Dado que la prevalencia depende de tantos factores no relacionados directamente con la causa de la enfermedad, en general los estudios de prevalencia no proporcionan pruebas claras de causalidad, aunque a veces puedan sugerirla. Sin embargo, son útiles para valorar la necesidad de asistencia sanitaria, planificar los servicios de salud o estimar las necesidades asistenciales.

Anteriormente, era común el cálculo de la llamada prevalencia de periodo (o lápsica), que buscaba identificar el número total de personas que presentaban la enfermedad o atributo a lo largo de un periodo determinado, combinando casos existentes con los nuevos que se detectaran durante el periodo de estudio. Dado que esta medida combina casos prevalentes con casos nuevos, esta medida es cada vez menos empleada, y es mejor no utilizarla.

Cuando se trata de una enfermedad rara y la población se encuentra en equilibrio, es decir, cuando la incidencia se mantiene constante y la duración de la enfermedad (la sobrevida) no ha cambiado, entonces se puede establecer una relación directa entre prevalencia y la incidencia:

$$\frac{p}{1-p} = I \times D$$

donde $p$ = prevalencia
donde $I$ = incidencia
donde $D$ = duración de la enfermedad

si la enfermedad es rara

$$p = I \times D \text{ ya que } (1 - p) \approx 1$$

## Incidencia

En los estudios epidemiológicos cuyo propósito es la investigación causal o la evaluación de medidas preventivas, el interés está dirigido a la medición del flujo que se establece entre la salud y la enfermedad, es decir, a la aparición de casos nuevos. Como ya se mencionó anteriormente, la medida epidemiológica que mejor expresa este cambio de estado es la incidencia, la cual indica la frecuencia con que ocurren nuevos eventos por unidad de tiem-

po. A diferencia de los estudios de prevalencia, en los que los sujetos de estudio se miden en un solo punto en el tiempo, la medición de incidencia implica al menos dos mediciones. Los estudios de incidencia inician con poblaciones de individuos susceptibles, libres del evento, en las que se observa la presentación de casos nuevos a lo largo de un periodo de seguimiento. De esta manera, los resultados no sólo indican el volumen final de casos nuevos aparecidos durante el seguimiento, sino que permiten estudiar la velocidad con que ocurren y estimar relaciones entre determinadas características de la población y enfermedades específicas. La incidencia de una enfermedad puede medirse de dos formas: mediante la tasa de incidencia (basada en el tiempo-persona) y mediante la incidencia acumulada (basada en el número de personas en riesgo). La tasa de incidencia (también denominada densidad de incidencia) expresa la ocurrencia de la enfermedad entre la población en relación con unidades de tiempo-persona, por lo que mide la velocidad de ocurrencia de la enfermedad. La incidencia acumulada, en cambio, expresa únicamente el volumen de casos nuevos ocurridos en una población durante un periodo, y mide la probabilidad de que un individuo desarrolle el evento en estudio. A la incidencia acumulada, por esta razón, también se le denomina riesgo.



*Tasa de incidencia o densidad de incidencia (TI).* Principal medida de frecuencia de enfermedad. Se define como "el potencial instantáneo de cambio en el estado de salud por unidad de tiempo, durante un periodo específico, en relación con el tamaño de la población susceptible en el mismo periodo". Para que una persona se considere en riesgo en el periodo de observación, éste debe iniciar sin tener la enfermedad (el evento en estudio) y ser susceptible de padecer la enfermedad.

El cálculo del denominador de la TI se realiza sumando los tiempos libres de enfermedad, es decir los tiempos en riesgo de padecer el evento, de cada uno de los individuos que conforman el grupo de estudio. Este número se puede expresar en años, meses, semanas o días, dependiendo del evento en estudio y se conoce como tiempo en riesgo o tiempo-persona.

El número de individuos que pasan del estado sano al estado enfermo durante cualquier periodo depende de tres factores: a) del tamaño de la población, b) de la amplitud del periodo de tiempo, y c) del poder patógeno de la exposición sobre la población. La tasa de incidencia mide este poder, y se obtiene dividiendo el número observado de casos entre el tiempo total en el que la población ha estado en riesgo, equivalente a la sumatoria de los periodos individuales en riesgo. Al sumar periodos de observación, que pueden variar de uno a otro individuo y considerar sólo el tiempo efectivo en riesgo, la TI corrige el efecto de entrada y salida de individuos del grupo en riesgo durante el periodo de seguimiento.

La TI no es una proporción —como la prevalencia y la incidencia acumulada—, dado que el denominador expresa unidades de tiempo y, en consecuencia, mide casos por unidad de tiempo. Esto hace que la magnitud de la TI no pueda ser inferior a cero ni tenga límite superior. La fórmula general para el cálculo de la TI es la siguiente:

$$\text{Tasa de incidencia} = \frac{\text{número de casos nuevos}}{\begin{array}{c}\text{suma de todos los periodos en riesgo durante el periodo definido en el estudio} \\ \text{(tiempo-persona)}\end{array}}$$

*Incidencia acumulada (IA).* Puede definirse como la probabilidad de desarrollar el evento, es decir, la proporción de individuos de una población que, en teoría, desarrollarían una enfermedad, si todos sus miembros fuesen susceptibles a ella y ninguno falleciese a causa de otras enfermedades. También se ha definido

simplemente como la probabilidad, o riesgo medio de los miembros de una población, de contraer una enfermedad en un periodo específico. Estima el riesgo promedio, es decir, la probabilidad "promedio" de que un miembro de la cohorte sufra el evento en un tiempo especifico.

Las cifras obtenidas mediante el cálculo de la IA son relativamente fáciles de interpretar y proporcionan una medida sumamente útil para comparar los diferentes riesgos de distintas poblaciones. Para calcular la IA, en el numerador se coloca el número de personas que desarrollan la enfermedad durante el periodo de estudio (llamados casos nuevos) y en el denominador el número de individuos en riesgo de desarrollar la enfermedad al comienzo del periodo de estudio. Se asume que todos los individuos completaron el periodo de estudio, es decir que en todos se tiene una segunda medición final. La incidencia acumulada es una proporción y, en consecuencia, sus valores sólo pueden variar entre 0 y 1. A diferencia de la tasa de incidencia, la IA es adimensional, pero siempre se expresa en relación con el tiempo de seguimiento. Su fórmula es la siguiente:

$$\text{IA} = \frac{\text{número de personas que contraen la enfermedad en un periodo determinado}}{\text{número de personas libres de la enfermedad en la población expuesta al riesgo en el inicio del estudio}}$$

Como la duración del periodo de observación influye directamente sobre la IA, su amplitud debe considerarse siempre que se interprete esta medida. Cuando los miembros de una población tienen diferentes periodos bajo riesgo—debido a que se incorporan o abandonan el grupo a lo largo del periodo de seguimiento—, la IA no puede calcularse directamente y su estimación deberá tomar en cuenta las pérdidas de seguimiento. Para esto se recomienda usar modelos de sobrevida o el método de Kaplan-Meier. Estos métodos estiman la probabilidad para cada evento en el momento en que ocurre. El denominador es la población en riesgo al momento en que ocurrió el evento.



## MEDIDAS DE ASOCIACIÓN

Las medidas de asociación son indicadores epidemiológicos que evalúan la fuerza con la que una determinada enfermedad o evento de salud (que se presume como efecto) se asocia con un determinado factor (que se presume como su causa).

Epidemiológicamente, las medidas de asociación son comparaciones de incidencias: la incidencia de la enfermedad en las personas que se expusieron al factor estudiado (o incidencia entre los expuestos) se compara con la incidencia de la enfermedad en las personas que no se expusieron al factor estudiado (o incidencia entre los no expuestos). Estadísticamente, estos indicadores miden la magnitud de la diferencia observada. Debido a que las medidas de asociación establecen la fuerza con la que la exposición se asocia con la enfermedad, bajo ciertas circunstancias estas medidas permiten realizar inferencias causales, especialmente cuando se evalúan en el contexto de un ensayo controlado. En este documento se abordará el cálculo de medidas de asociación para variables dicotómicas.

Las medidas de asociación más sólidas se calculan utilizando la incidencia, ya que esta medida de frecuencia nos permite establecer, con mayor certeza, que el efecto (el evento o enfermedad) es posterior a la causa (la exposición). En estos casos, se dice, existe una correcta relación temporal entre la causa y el efecto. Sin embargo, hay estudios en los que no es posible

calcular la incidencia, debido a que no existe la información suficiente (como las encuestas transversales y la mayoría de los estudios de casos y controles). En estos casos puede estimarse la asociación entre el evento y la exposición al comparar las prevalencias mediante la razón de prevalencias (RP), o los momios de desarrollar la enfermedad dada la exposición, mediante la razón de momios (RM).

En general, hay dos tipos de medidas de asociación: las de diferencia (o de efecto absoluto) y las de razón (o de efecto relativo).

## MEDIDAS DE DIFERENCIA

Como su nombre lo indica, estas medidas expresan la diferencia existente en una misma medida de frecuencia (idealmente la incidencia) entre dos poblaciones. La más importante de ellas es el riesgo atribuible.

### Riesgo atribuible

En general, las medidas de diferencia indican la contribución de un determinado factor en la producción de enfermedad entre los que están expuestos a él. Su uso se basa en la suposición de que tal factor es responsable de la aparición de un exceso o déficit de casos de la enfermedad (según lo plantee la hipótesis), y en la presunción de que, de no comprobarse la hipótesis, los riesgos de padecer la enfermedad en ambos grupos serían equivalentes. Por este motivo, se dice que las medidas de diferencia indican el riesgo de enfermar que podría evitarse si se eliminara la exposición. Como sinónimo se emplea el término riesgo atribuible. Estas medidas se calculan de la siguiente manera:

$$\text{Diferencia} = E_i - E_o \times 100$$

donde,

$E_i$ = es la frecuencia de enfermar o morir de un grupo expuesto, y

$E_o$ = es la frecuencia de enfermar o morir en el grupo no expuesto.

El resultado se interpreta de la siguiente forma:

Valor = 0 indica no asociación (valor nulo).

Valores < 0 indica asociación negativa y puede tomar valores negativos hasta infinito.

Valores > 0 indica asociación positiva y puede tomar valores positivos hasta infinito.

Debe señalarse que el término *riesgo atribuible* carece de justificación cuando no existe una relación causa-efecto entre la exposición y la enfermedad. No obstante, como la diferencia de incidencias —ya sea diferencia de tasas de incidencia (DTI) o diferencia de riesgos (DR)— puede, en el contexto de ensayos controlados, llegar a indicar diferencias verdaderamente atribuibles a la exposición, estas medidas se siguen usando para estimar la magnitud de problemas de salud pública, aunque se usan poco en estudios observacionales.

La diferencia de prevalencias (DP), usada en estudios transversales, puede ser en algunas condiciones un estimador aceptable de la diferencia de incidencias, pero sus resultados sólo indican asociación y no causalidad y deben ser interpretados con cautela.

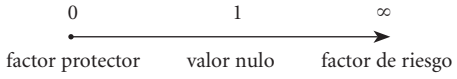
## MEDIDAS DE RAZÓN

Estas medidas también cuantifican las discrepancias en la ocurrencia de enfermedad en grupos que difieren en la presencia o no de cierta característica. Como se señaló antes, una razón puede calcularse tanto para dos eventos en una misma población como para un solo evento en dos poblaciones. Las razones que con mayor frecuencia se calculan son del segundo tipo, y se obtienen con la siguiente fórmula:

$$\text{Razón} = \frac{\text{medida de frecuencia en un grupo expuesto } (E_i)}{\text{medida de frecuencia en un grupo no expuesto } (E_o)}$$

La razón representa cuántas veces más (o menos) ocurrirá el evento en el grupo expuesto al factor, comparado con el grupo no expuesto. El resultado se interpreta de la siguiente forma:

- Valor = 1 indica ausencia de asociación, no–asociación o valor nulo.
- Valor < 1 indica asociación negativa, factor protector.
- Valor >1 indica asociación positiva, factor de riesgo.




Conforme el resultado se aleja más de la unidad, la asociación entre el factor y la enfermedad es más fuerte. Un valor de 4 indica que el riesgo de enfermar entre los expuestos es cuatro veces mayor que entre los no expuestos. Asimismo, un valor de 0.25 indicaría que el riesgo de enfermar entre los expuestos es cuatro veces menor que entre los no expuestos.

La incidencia es una de las medidas de frecuencia que más se emplean en la construcción de las medidas de razón. Con la densidad de incidencia se obtiene la razón de densidad de incidencia (RDI), y con la incidencia acumulada se obtiene la razón de incidencia acumulada (RIA), también llamada riesgo relativo (RR). Ambas medidas —que se obtienen en estudios de cohorte— permiten asumir inferencia etiológica, ya que siempre implican la posibilidad de establecer adecuadamente una relación de temporalidad causal.

### Razón de densidad de incidencia

La razón de densidad de incidencia (RDI) es útil para determinar la velocidad con la que se pasa de un estado a otro, por ejemplo de sano a enfermo, según se está o no expuesto a una determinada condición, misma que se hipotetiza como una probable causa de este paso de sano a enfermo. Se utiliza especialmente cuando el periodo de inducción o latencia de una enfermedad es largo; cuando la enfermedad no es muy rara, y cuando hay razones para suponer que existirán pérdidas importantes durante el seguimiento.

$$\text{RDI} = \frac{\text{densidad de incidencia en expuestos}}{\text{densidad de incidencia en no expuestos}}$$

$$\text{RDI} = \frac{DI_1}{DI_0} = \frac{I_1/TP_1}{I_0/TP_0}$$

donde:

- $DI_1$ = Densidad de incidencia (o tasa de incidencia) en expuestos
- $DI_0$ = Densidad de incidencia (o tasa de incidencia) en no expuestos
- $I_1/TP_1$ = casos incidentes / tiempo persona de seguimiento en expuestos
- $I_0/TP_0$ = casos incidentes/ tiempo persona de seguimiento en no expuestos

### Razón de incidencia acumulada o riesgo relativo

Compara el riesgo de enfermar del grupo de expuestos ($LA_i$) con el riesgo de enfermar del grupo de no expuestos ($LA_o$). Es útil, si lo que se desea es conocer la probabilidad de padecer la enfermedad en función de la exposición, y es la medida que mejor refleja su asociación.

$$RR = \frac{\text{incidencia acumulada (o riesgo) en expuestos}}{\text{incidencia acumulada (o riesgo) en no expuestos}}$$

$$RDI = \frac{IA_i}{IA_o} = \frac{a/n_i}{c/n_o}$$

donde,

$IA_i$ es la incidencia acumulada o riesgo de enfermar entre los expuestos, e

$IA_o$ es la incidencia acumulada o riesgo de enfermar entre los no expuestos (para observar gráficamente la ubicación de las celdas a, c, $n_i$ y $n_o$, véase la siguiente tabla de 2 × 2).

Formato estándar de una tabla de 2 × 2

| | Exposición: presente | Exposición: ausente | |
|---|---|---|---|
| casos | a | b | total de casos ($n_i$) |
| no casos | c | d | total de no casos ($n_o$) |
| | total de expuestos ($m_i$) | total de no expuestos ($m_o$) | total de sujetos (n) |

### Razón de prevalencias

La razón de prevalencias (RP) se utiliza en los estudios transversales y se calcula de forma similar a la estimación del RR en los estudios de cohorte. Si la duración del evento que se estudia es igual para expuestos y no expuestos, la RP puede ser un estimador de la RDI, pero en general esta medida subestima la RDI.

$$RP = \frac{\text{prevalencia en expuestos}}{\text{prevalencia en no expuestos}}$$

$$P = \frac{P_1}{P_o} = \frac{a/n_1}{c/n_o}$$

donde,

$P_1$ es la prevalencia de la enfermedad en los expuestos, y

$P_o$ es la prevalencia de la enfermedad en los no expuestos (de acuerdo con la ubicación de las celdas a, c, $n_i$ y $n_o$ que se obtiene del formato estándar de una tabla de 2 x 2).

### Razón de momios

La razón de momios (RM) es una medida utilizada en los estudios de casos y controles —donde los sujetos son elegidos según la presencia (casos) o ausencia de enfermedad (controles), desconociéndose el tamaño de la población de donde provienen—, por lo que no es posible calcular la incidencia de la enfermedad. La RM también se conoce con los términos en inglés *odds ratio* (OR) y en español *razón de ventaja, razón de disparidad* o *razón de productos cruzados* (RPC). La RM es un buen estimador de la RDI, sobre todo cuando los controles son representativos de la población de la que han sido seleccionados los casos. La RM también puede ser un buen estimador del RR, cuando la enfermedad es rara. Esta medida se calcula por medio de la obtención del cociente de los productos cruzados de una tabla tetracórica:

$$RM = \frac{a/c}{b/d} = \frac{ad}{bc}$$

donde,

Exposición

| | presente | ausente | |
|---|---|---|---|
| casos | a | B | Total de casos ($n_i$) |
| controles | c | D | Total de controles ($n_o$) |
| | Total de expuestos ($m_i$) | Total de expuestos ($m_o$) | Total de sujetos (n) |

Al igual que en las medidas anteriores, esta fórmula expresa el caso más sencillo, cuando la exposición y la enfermedad se reportan simplemente como presentes o ausentes. El resultado se interpreta de la misma forma que en el resto de las medidas de razón. Cuando la RM tiene un valor de 1 (o nulo), el comportamiento del factor es indiferente; si el valor es superior a 1, el factor puede considerarse de riesgo, y si es inferior a 1, se valora como factor protector.

## MEDIDAS DE IMPACTO

La razón de densidad de incidencia, el riesgo relativo y la razón de momios describen la asociación entre la exposición y el evento en términos de la magnitud de la fuerza de su asociación, dato importante cuando evaluamos la existencia de asociaciones causales. Sin embargo, estas medidas no pueden traducirse fácilmente en el contexto de la salud de la población. ¿Qué tan importante es una exposición? ¿Qué proporción de las enfermedades puede atribuírsele a esta variable? ¿Qué impacto tendría en la población controlar esa exposición? Para poder estimar el efecto de cierta exposición en la población en estudio o en la población blanco, se requiere estimar otro tipo de medidas, conocidas como medidas de impacto.

Las principales medidas de impacto potencial son el riesgo atribuible (o fracción etiológica), que se estima cuando el factor de exposición produce un incremento en el riesgo (RR $>$ 1), y la fracción prevenible, relacionada con factores que producen una disminución en el riesgo (RR $<$ 1).

## RIESGO ATRIBUIBLE

Anteriormente, era muy común el uso del término *fracción etiológica* para referirse a este indicador; sin embargo, en la actualidad se recomienda utilizarlo únicamente para referirse a relaciones causales bien demostradas. El término más usual y conservador es el riesgo atribuible proporcional. Para esta última medida, se han derivado dos dimensiones, el *riesgo atribuible proporcional en el grupo expuesto* ($RAP_{Exp}$) y el *riesgo atribuible proporcional en*

*la población blanco* (RAPP). Ambas medidas son proporciones, por lo que toman valores entre cero y uno e indican la importancia relativa de la exposición al factor en estudio en relación con el total de eventos. El $RAP_{Exp}$ tiene interpretación en el ámbito de la población en estudio, mientras que el RAPP expresa la importancia en el ámbito poblacional, o población blanco.

El $RAP_{Exp}$ estima la proporción de eventos en el grupo expuesto que pueden atribuirse a la presencia del factor de exposición. En otras palabras, refleja el efecto que podría esperarse en el grupo expuesto de la población en estudio, si se eliminara el factor de riesgo en cuestión. El $RAP_{Exp}$ puede calcularse, utilizando la siguiente fórmula:

$$RAP_{Exp} = \frac{DI_E - DI_{NE}}{DI_E} = \frac{RDI-1}{RDI}$$

donde

$DI_E$ = densidad de incidencia en expuestos,
$DI_{NE}$ = densidad de incidencia en no expuestos, y
RDI = razón de densidad de incidencia

El $RAP_{Exp}$ puede estimarse también en estudios donde la medida de frecuencia es la incidencia acumulada, utilizando el riesgo relativo. Además, dado que la razón de momios es un buen estimador de la RDI, el $RAP_{Exp}$ también puede estimarse en los estudios de casos y controles, mediante la siguiente fórmula:

$$RAP_{Exp} = \frac{RM-1}{RM}$$

Para ilustrar su interpretación y cálculo, supongamos que desea estimarse el $RAP_{Exp}$ de los resultados derivados de un estudio de casos y controles sobre tabaquismo y cáncer pulmonar. En el mencionado estudio, se documenta una asociación entre el riesgo de cáncer de pulmón y el tabaquismo (RM) de 12.5. El $RAP_{Exp}$ podría estimarse dividiendo 12.5 menos 1 entre 12.5, lo que daría un $RAP_{Exp}$ de 0.92 (o 92%), lo que indicaría que el 92 % de los casos de cáncer pulmonar en el grupo expuesto al tabaco podrían atribuirse a esta exposición. Esto significa que el $RAP_{Exp}$ indica el porcentaje de casos en el grupo expuesto que podría prevenirse, si se eliminara la exposición, asumiendo que ésta es la única causa del evento y que el resto de las causas de cáncer de pulmón se distribuyen de igual manera entre los fumadores (grupo expuesto) y los no fumadores (grupo no expuesto), como se indica en la figura 1. En el ejemplo anterior indicaría que se podrían prevenir alrededor de 92% de los casos de cáncer de pulmón que ocurren en el grupo de fumadores.

El RAPP puede considerarse como una proyección del $RAP_{Exp}$ hacia la población total. En este caso, los resultados obtenidos en el grupo de expuestos se extrapolan hacia la población blanco, estimando el impacto de la exposición en el ámbito poblacional. Siguiendo el ejemplo anterior, la estimación del RAPP nos indicaría cuántos casos de cáncer de pulmón en la población total son atribuibles al tabaco o podrían evitarse, suponiendo que se eliminara el tabaquismo en la población general. EL RAPP se estima ponderando el $RAP_{Exp}$ de acuerdo con la proporción de sujetos expuestos en la población blanco. El RAPP puede estimarse con la siguiente fórmula:

$$RAPP = \frac{P_{exp}\,(RDI-1)}{P_{exp}\,(RDI-1) + 1}$$

Al igual que en el caso anterior, el RAPP puede estimarse en estudios de cohorte, donde se estima la incidencia acumulada, o en estudios de casos y controles, donde se estima la razón de momios. En este último caso, puede utilizarse la prevalencia de exposición en los controles para estimar la prevalencia en la población blanco o población de referencia. En

el estudio antes mencionado sobre tabaquismo y cáncer pulmonar, se observó una prevalencia de 28.5 de tabaquismo en el grupo control. Dado que la serie de controles puede considerarse como representativa de la población base, en este estudio podría estimarse directamente el RAPP, lo que daría una fracción de 0.76. Esta última cifra indicaría que, en la población blanco, 76% de los casos de cáncer pulmonar pueden atribuírsele al tabaquismo, asumiendo que ésta es su única causa. Mediante el cálculo del $RAP_{Exp}$ y del RAPP es posible identificar diversos escenarios:

a) con un RR alto y una prevalencia de expuestos alta, la reducción del riesgo de enfermedad puede considerarse como de alto impacto;
b) cuando el RR es bajo y la prevalencia de expuestos es alta, la supresión del factor de riesgo posee un impacto moderado, pero notable entre los expuestos;
c) cuando el RR es alto pero la prevalencia de expuestos es baja, la eliminación del factor de riesgo tiene un impacto relativamente bajo, tanto entre la población blanco como entre los expuestos, y
d) cuando el RR es bajo y la prevalencia de expuestos también es baja, la eliminación del factor de riesgo no es una prioridad en salud pública, ya que su impacto en la población blanco y en los expuestos sería irrelevante.

### Fracción prevenible

Esta medida se aplica cuando se obtienen factores protectores o negativos (RR < 1) a partir de las medidas de asociación. También existen dos modalidades: fracción prevenible poblacional y fracción prevenible entre expuestos.



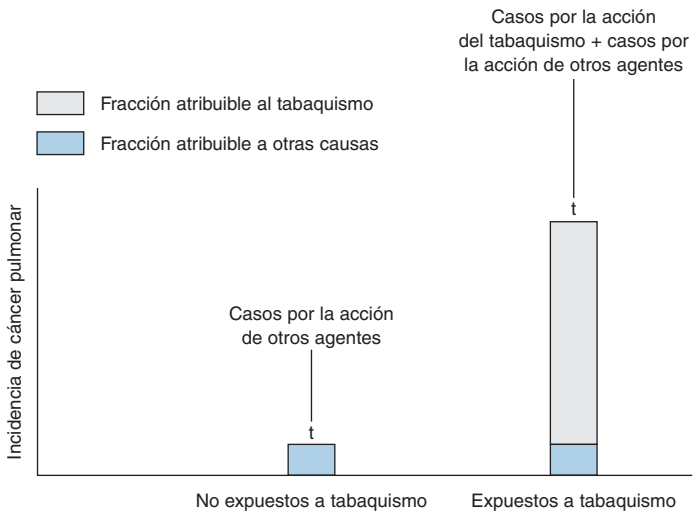



**Figura 1** **Representación hipotética de un estudio de cohorte para evaluar el efecto del tabaquismo sobre riesgo de desarrollar cáncer de pulmón.**

La fracción prevenible poblacional es la proporción de todos los casos nuevos que potencialmente podrían haber ocurrido entre la población general en un determinado periodo, en ausencia de una exposición protectora específica, o bien, la proporción de casos potenciales que serían realmente prevenibles o evitados por un agente protector si existiera la exposición entre la población.

Finalmente, la fracción prevenible para los expuestos es la proporción de casos nuevos entre los expuestos que potencialmente podrían presentarse en un determinado periodo, en ausencia de una exposición protectora particular (por ejemplo, una vacuna). Es decir, es la proporción de casos expuestos potenciales que realmente se evitarían en el caso de que la población se expusiera al factor protector.

## BIBLIOGRAFÍA

Argimon Pallás JM, Jiménez Villa J. Métodos de investigación clínica y epidemiológica. Barcelona: Harcourt, 2000.

Gordis L. Epidemiology. Philadelphia: W.B. Saunders Co., 1995.

Jenicek M. Epidemiología: la lógica de la medicina moderna. Barcelona: Masson, 1996.

Kleimbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Nueva York: Van Nostrand Reinhold Co., 1982.

Kleinbaum DG, Sullivan KM, Barker ND. ActivEpi companion textbook: A supplement for use with the ActivEpi CD-ROM. Nueva York: Springer-Verlag, 2003.

Martínez NF, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V. Salud pública. Madrid: McGraw-Hill-Interamericana, 1998.

Rothman JK. Modern epidemiology. Boston: Little Brown & Co., 1986.

# IV Estudios clínicos experimentales*

*Juan José Calva Mercado*

Los estudios en los cuales la aleatorización es utilizada para la asignación de sujetos se denominan *experimentos.*[1]

Mientras que muchos profesionales de la salud consideran que los epidemiólogos no realizan "verdaderos" experimentos, una tradición bien definida de investigación experimental ha crecido en la epidemiología, en gran medida, en respuesta a la necesidad de probar la eficacia de nuevos medicamentos y nuevas vacunas.

Los estudios experimentales se utilizan en las investigaciones realizadas en grupos y colectividades tanto para estudiar las causas de las enfermedades (investigación etiológica) como para evaluar las intervenciones de programas de salud (investigación evaluativa).

Mientras que el uso de estudios experimentales en la investigación causal está muy limitado por motivos éticos, legales y prácticos, ya que se trabaja en poblaciones humanas, su empleo es común en la investigación evaluativa para estudiar la eficacia y efectividad de intervenciones médicas y sanitarias. Su uso en la evaluación de acciones sanitarias es más factible. Se usan tanto en investigaciones clínicas, para probar la eficacia de un nuevo medicamento o una medida preventiva (ensayos clínicos), como en investigaciones comunitarias para evaluar la efectividad de un programa de intervención determinado.

La esencia de los estudios experimentales es que el investigador decide cuáles individuos serán sometidos a la intervención (grupo experimental) y quienes estarán en el grupo comparativo (grupo control o de contraste). En otras palabras, el investigador manipula a su voluntad las variables independientes (exposición o intervención) y los sujetos son asignados al grupo experimental o de control. Es justamente esta propiedad la que distingue a los ensayos clínicos experimentales de los estudios observacionales (de cohortes), ya que en éstos el propio paciente o su médico tratante decide quién se somete (y quién no) a la maniobra en evaluación; esta decisión obedece a múltiples razones, algunas estrechamente ligadas al pronóstico de la enfermedad. Debido a que los ensayos clínicos controlados son estudios diseñados con antelación, la asignación de la maniobra experimental por el investigador puede seguir diversos procedimientos; cuando es mediante un sorteo el estudio se conoce como un ensayo clínico aleatorizado, o controlado por sorteo.

Cabe aclarar la connotación de diversos términos empleados en la anterior definición y a lo largo del presente capítulo. El uso del término tratamiento (o terapia) tiene un sentido amplio; se refiere no sólo a un medicamento, también incluye otro tipo de intervenciones (o maniobras) tales como un procedimiento quirúrgico, una medida preventiva (o profiláctica), un programa educativo, un régimen dietético, etcétera. De igual manera, el término evento (o desenlace) se puede referir a una diversidad de resultados, tales como: mediciones bioquímicas, fisiológicas o microbiológicas; eventos clínicos (intensidad del dolor, aparición de infecciones oportunistas, desarrollo de un infarto al miocardio, recaída de una

leucemia aguda, etc.); escalas de actividad de una enfermedad (como la del lupus eritematoso generalizado); mediciones de bienestar o funcionalidad (calificación de Karnofsky, escala de calidad de vida) o el tiempo de supervivencia. Por último, el grupo control se refiere al grupo de individuos que recibe una intervención que sirve de contraste para evaluar la utilidad relativa de la terapia experimental y que no necesariamente tiene que ser un placebo, pues en ocasiones lo más adecuado (y ético) es que sea el tratamiento estándar, es decir, la mejor alternativa terapéutica vigente en el momento del diseño del experimento clínico.

Los estudios experimentales constituyen el diseño más completo de investigación epidemiológica y proporcionan una firme evidencia de causalidad.

## ¿POR QUÉ Y CUÁNDO SON NECESARIOS?

La pregunta nos remite a reflexionar sobre dónde surge la idea de que un cierto tratamiento pueda modificar la historia natural de una enfermedad y acerca de la necesidad de contar con suficientes observaciones sistematizadas para conocer el verdadero efecto de la terapia en cuestión en los seres humanos, antes de prescribirla de manera rutinaria a los enfermos.



En ocasiones, surge la idea de que un fármaco pudiera ser clínicamente útil al comprender su íntimo mecanismo de acción, así como la patogenia a nivel celular y molecular de una determinada enfermedad. De hecho, en ocasiones resulta muy tentador intentar predecir el efecto de un cierto medicamento con base únicamente en la información generada en el laboratorio, tanto *in vitro* como en modelos animales de experimentación. Sin embargo, por muy profundo que sea este conocimiento siempre será incompleto si no se tiene la experiencia en seres humanos intactos; de no ser así, se corre el riesgo de sorpresas desagradables. Por ejemplo, se sabe que el antimetabolito citarabina interfiere con la síntesis de la pirimidina y que es capaz de inhibir *in vitro* a varios virus con ADN, incluyendo al virus herpes varicela-zóster. A algún ingenuo clínico se le ocurriría que pudiera beneficiar con este medicamento a sus enfermos con herpes zóster generalizado; afortunadamente a un grupo de médicos escépticos se les ocurrió comparar la evolución de un grupo de estos enfermos, a quienes se les administró la citarabina, con la de otro grupo de enfermos semejantes, quienes sólo recibieron placebo, y demostraron no sólo lo ineficaz del antiviral sino una peor evolución en los enfermos que la habían recibido, explicable por sus efectos inmunosupresores.[2]

Otras dos fuentes de ideas sobre el posible valor de una cierta terapia suelen ser: 1) las observaciones empíricas de clínicos perspicaces; así, por serendipia surgió la idea de que la amantadina pudiera ser de utilidad en los enfermos de Parkinson,[3] puesto que a quienes se les prescribía para prevenir la influenza mostraban mejoría de sus manifestaciones neurológicas, o como el caso de la reducción de crisis de fiebre familiar del Mediterráneo con el uso de colchicina administrada para prevenir ataques de gota (en estos ejemplos, el valor de estos tratamientos no provino de un entendimiento de la patogenia de estas enfermedades, la que de hecho, aún no se conoce bien); y 2) las observaciones de estudios poblacionales don-

de se llega a establecer una asociación directa o inversamente proporcional entre la frecuencia de una enfermedad y alguna condición ambiental, como ha sido el caso de una relativa menor frecuencia de enfermedad coronaria en poblaciones con mayor ingesta de alimentos ricos en antioxidantes o el de cáncer de colon y una dieta con alto contenido en fibra.

El punto esencial es que cualquiera que sea el origen de las hipótesis del posible beneficio de una terapia, éstas deben probarse, y demostrarse como ciertas, mediante estudios clínicos; es decir, mediante la observación sistematizada y objetiva del efecto de la terapia en seres humanos que la reciben, y su comparación con lo que habitualmente sucede en un grupo de enfermos sin ella.

Hay circunstancias en las que, por una parte, se conoce muy bien la historia natural de una enfermedad y que es tan consistente que es posible, con razonable certeza, predecir su curso clínico (generalmente muy desfavorable) y, por la otra, que el beneficio de una cierta terapia en este tipo de enfermedades es tan dramático e incuestionable que resulta innecesario realizar un ensayo clínico para aceptar su uso generalizado. Es decir, es suficiente comparar la evolución clínica de un grupo de casos tratados con la nueva terapia con lo que habitualmente pasa en los enfermos antes del acceso a ésta (controles históricos), tal y como ha sucedido con el beneficio de los antibióticos en el tratamiento de la neumonía bacteriana, de los antifímicos en la meningitis tuberculosa o de la cirugía abdominal en la apendicitis: no hay alguien en el mundo que haya exigido realizar un ensayo clínico en el que un grupo de estos enfermos recibieran un placebo. Sin embargo, ésta es una circunstancia excepcional; lo común es que diferentes individuos con la "misma" enfermedad tengan cursos disímiles (en ocasiones, en sentidos totalmente opuestos) e impredecibles, de tal suerte que esta incertidumbre pronóstica impide deslindar qué tanto la mejoría o las complicaciones observadas en un grupo de enfermos que reciben una nueva terapia es atribuible a ésta, o bien, es parte de su historia natural o consecuencia de otros determinantes ajenos a la intervención en evaluación. Por ejemplo, un grupo de enfermos puede tener una mejor evolución clínica a pesar de recibir un tratamiento totalmente inútil, porque son pacientes en etapas tempranas de su enfermedad, reciben otros medicamentos concomitantes o mejor atención médica, están bajo el efecto placebo de la terapia novedosa o porque comparten el entusiasmo de los investigadores. Es en estos casos, que constituyen más la regla que la excepción, cuando es indispensable (ética y científicamente) realizar (y exigir) el diseño, la conducción y la publicación formal de un estudio clínico experimental; el ensayo clínico controlado por sorteo es el diseño metodológico más confiable para distinguir si el beneficio atribuible a un tratamiento es real o sólo un espejismo.

En la historia de la medicina abundan los bochornosos ejemplos de tratamientos (novedosos en su época) a los que se les atribuyeron espectaculares beneficios (por lo que multitudes los recibieron) y que no fue hasta que un ensayo clínico aleatorizado los puso en su justo lugar: en el archivo de terapias inútiles e, incluso, dañinas. Tal es el caso de la congelación gástrica en el tratamiento de la úlcera péptica,[4] de la ligadura de la arteria mamaria interna en la prevención de la angina de pecho recurrente, de los esteroides en los pacientes con sepsis grave,[5] de la plasmaféresis en los pacientes con polimiositis,[6] entre muchos otros ejemplos.

También hay ejemplos de intervenciones terapéuticas o profilácticas ampliamente arraigadas en la práctica médica cotidiana actual pero cuyo beneficio real ha sido, en mayor o menor grado, cuestionado; tal como sucede con la vacuna antituberculosa con el ba-

cilo de Calmette-Guerin, el uso de antibióticos en la otitis media aguda no complicada en los niños[7] o el internamiento de enfermos en una unidad coronaria. Sin embargo, éstos son ejemplos en los que resulta casi imposible (por razones éticas o logísticas) llevar a cabo en la actualidad un estudio clínico experimental, por lo que parece que estamos condenados a no resolver el dilema de su verdadero costo-beneficio. De ahí que hay quien recomienda que se realicen los ensayos clínicos controlados de manera temprana en la fase de evaluación de nuevas terapias, antes de que sean adoptadas de manera irreflexiva por la comunidad médica. Sin embargo, también debe tomarse en cuenta que lo común es que los ensayos clínicos sean estudios costosos y logísticamente complejos, por lo que deben plantearse teniendo información completa de su farmacología, de su toxicidad y preliminar de su posible eficacia.



## Tipos de estudios clínicos experimentales (ECE)

Se pueden identificar tres tipos de ECE: los experimentos de laboratorio, los ensayos clínicos y las intervenciones comunitarias. La clasificación depende de la duración aproximada del estudio y de la selección de los sujetos.[1]

Los experimentos de laboratorio son de corta duración —horas o días— y, por lo tanto, se utilizan para determinar o estimar respuestas biológicas y/o de comportamiento agudas que se cree son debidas a los factores de riesgo de la enfermedad. Típicamente, la población de estudio está muy restringida o auto-seleccionada y esto condiciona que rara vez es representativa de la población blanco.

El ensayo clínico es de mayor duración —días hasta años— y habitualmente está restringido a poblaciones muy seleccionadas, como a un grupo de sujetos tamizados, casos diagnosticados u otros voluntarios. El objetivo princi-

**Cuadro I Características de los diferentes tipo de estudios clínicos experimentales***

| *Tipo de estudio* | *Objetivos* | *Duración habitual* |
|---|---|---|
| Laboratorio | Prueba hipótesis etiológicas y estima respuestas biológicas y/o de comportamiento agudas<br>Sugiere eficacia de una intervención para modificar factores de riesgo en una población | Horas o días |
| Ensayo clínico | Prueba hipótesis etiológicas y estima efectos de salud a largo plazo<br>Prueba eficacia de intervenciones que modifican el estado de salud<br>Sugiere factibilidad de intervenciones | Días hasta años |
| Intervención comunitaria | Identifica personas de “alto riesgo”<br>Prueba eficacia y efectividad de intervenciones clínicas/sociales que modifican el estado de salud dentro de poblaciones particulares<br>Sugiere políticas y programas de salud pública | No menos de 6 meses |

* Modificado de: Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. Principles and Quantitative Methods. New York, NY: John Wiley & Sons, INC, 1982, pág. 41.

pal es probar el efecto posible —por ejemplo, eficacia— de una intervención preventiva o terapéutica. Además, si el estudio es éticamente factible, el ensayo clínico puede ser utilizado para evaluar relaciones etiológicas específicas, teniendo un componente terapéutico o preventivo no inmediato.

Por último, la intervención comunitaria tiene larga duración —al menos 6 meses— y difiere de los dos tipos previos en que el estudio es realizado en poblaciones formadas naturalmente o pertenecientes a un contexto sociopolítico o socioeconómico particular. Por lo tanto, los objetivos de la intervención comunitaria comúnmente pretenden la implementación y evaluaciones dirigidas a la prevención primaria a través de la modificación de factores de riesgo en una población bien definida. En general, se desea determinar los beneficios potenciales de modificar ciertos comportamientos de los individuos, características biológicas o aspectos del medio ambiente.

En el cuadro I se describen en forma resumida las características de cada uno de los estudios.

## ¿QUÉ DETERMINA LA VALIDEZ DE SUS RESULTADOS?

Esta pregunta tiene que ver con la confianza que se llegue a tener de que los resultados del ensayo clínico revelen con exactitud la dirección y magnitud de lo que les pasa a los sujetos bajo estudio y que esto sea realmente atribuible al tratamiento en prueba. Es necesario que se tengan los elementos para evaluar si los resultados representan una estimativa no sesgada del efecto de la intervención o si, por el contrario, son conclusiones falsas determinadas por algún error sistemático. A continuación se describen los principales componentes, porque no son los únicos, que determinan la validez interna de un ensayo clínico.

### Asignación del tratamiento en prueba por medio de sorteo

La decisión de a cuál grupo pertenecerá cada uno de los individuos participantes en el estudio debe ser ajena a cualquier predisposición o prejuicio del investigador y la mejor manera de lograrlo es mediante el empleo de un proceso aleatorio: es únicamente el azar quien se encarga de asignar a los sujetos (figura 1).

En la década de los setenta y principios de los ochenta era común que a los pacientes con insuficiencia arterial cerebral se les practicara una cirugía de revascularización extra-intracraneana mediante la colocación de un puente en el que se unía una rama de la arteria carótida externa (la temporal superficial) con una rama cortical de la arteria carótida interna (la cerebral media). Los cirujanos que la practicaban entonces lo hacían con mucho entusiasmo convencidos de que su procedimiento realmente prevenía la oclusión de las arterias cerebrales en sus nobles pacientes. Esta creencia se basaba en la observación repetida de que grupos de pacientes así operados con menor frecuencia sufrían eventos isquémicos cerebrales en comparación con los enfermos que no eran sometidos a la cirugía. El detalle importante es que estas eran observaciones provenientes de la práctica clínica cotidiana, es decir, no se habían obtenido de un estudio planeado en donde un grupo de investigadores clínicos hubieran decidido, mediante un procedimiento semejante al de un "volado", quiénes era operados y quiénes no. En el quehacer clínico cotidiano no es nada raro que el cirujano ofrezca sus servicios (o que el paciente se deje operar) preferentemente a pacientes que están relativamente en mejores condiciones que otros; es decir, que desde antes de la

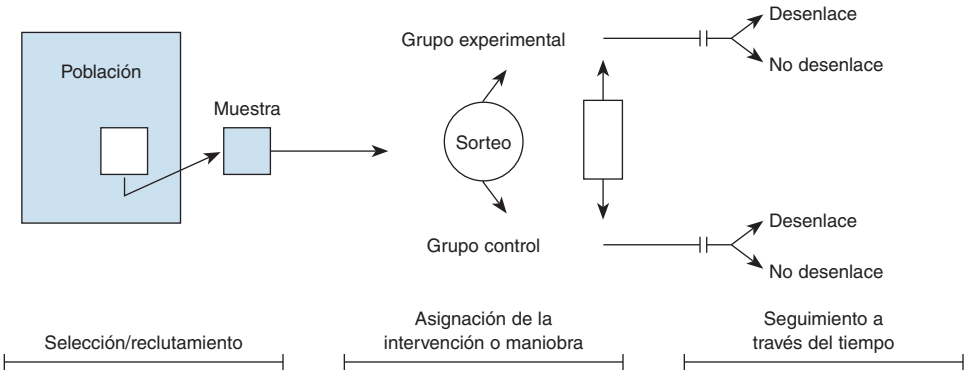



**Figura 1** **Diagrama de la estructura de un ensayo clínico controlado por sorteo (aleatorizado). La población de interés se define con base en los criterios de inclusión/exclusión y la muestra estudiada generalmente se selecciona y recluta por su accesibilidad. La asignación de los participantes al grupo experimental (quien recibe la intervención bajo evaluación) o al grupo control (quien recibe la intervención de contraste: el placebo o la maniobra habitual) la realiza el propio investigador mediante un procedimiento aleatorio. El efecto de la intervención en estudio se mide al comparar la incidencia del desenlace de interés en el grupo experimental con la del grupo control**



cirugía gozan de un inherente mejor pronóstico en cuanto al riesgo de oclusión arterial, de complicaciones de la cirugía e, incluso, de supervivencia; en cambio (y por razones totalmente entendibles), no suelen hacerlo con pacientes más deteriorados, en quienes hay menos posibilidades de que su intervención quirúrgica resulte exitosa. Esta práctica puede hacer entonces que, como los pacientes con menos posibilidades de una buena evolución son incluidos preferentemente en el grupo (cohorte) control (o de contraste), un procedimiento ineficaz parezca (de manera distorsionada) como todo lo contrario. Es justamente mediante el sorteo (también conocido como asignación aleatoria de la maniobra, que sólo puede hacerse en un estudio experimental) que se pretende que todos los pacientes tengan la misma oportunidad de recibir el tratamiento en prueba; de tal suerte que las diferencias en el desenlace clínico entre los sometidos (o no) al tratamiento sean explicables únicamente por este hecho y no por determinantes ajenos a la maniobra en estudio. Resulta interesante que fue mediante un enorme estudio experimental en humanos, aleatorizado, multicéntrico y publicado en 1987, que se logró demostrar tanto que a los pacientes así operados en realidad les va peor que a los tratados únicamente con medicamentos en el periodo posoperatorio inmediato como que la evolución clínica a largo plazo era idéntica entre los sometidos o no a la revascularización quirúrgica.[8]

La evolución clínica de los enfermos obedece a múltiples causas, el tratamiento sólo es una de ellas; de tal forma que es común que la gravedad de la enfermedad, la presencia de comorbilidad y toda una gama de determinantes pronósticos (unos que se pueden identificar y

otros muchos que se desconocen) hagan que se tenga una falsa impresión de las bondades o perjuicios de los tratamientos que se ofrecen. Es por ello que los médicos que toman decisiones basadas en información proveniente de estudios observacionales (o de experimentales no aleatorizados) están propensos a caer en este tipo de errores. De hecho, se ha documentado que los estudios en los que el tratamiento se ha asignado por cualquier otro método que no haya sido un sorteo tienden a mostrar efectos de mayor beneficio (y frecuentemente falsos) que lo observado en los ensayos clínicos aleatorizados. La ventaja de la aleatorización es que logra, si el tamaño de la muestra es suficientemente grande, que los determinantes de la evolución de los enfermos (tanto los conocidos como los desconocidos) estén distribuidos de manera equilibrada entre el grupo de enfermos que reciben la maniobra bajo estudio (grupo experimental) y los que no la reciben (grupo control).

**Seguimiento completo de los individuos en estudio**

Este punto tiene que ver tanto con que se logre un seguimiento completo de todos los sujetos participantes en el estudio durante el tiempo programado como con que se incluya en el análisis a todos los individuos respetando su pertenencia al grupo (experimental o control) al que fueron originalmente asignados.

Todo paciente que ingresa al ensayo debe ser tomado en cuenta en el análisis y conclusiones; de lo contrario, si un número importante de ellos se reporta como "sin seguimiento suficiente" la validez del estudio se verá seriamente cuestionada. Mientras más individuos se pierden mayor será la posibilidad de sesgo en el estudio debido a que los pacientes que no completan su seguimiento pueden tener un pronóstico diferente al de quienes sí permanecen hasta el final del estudio. Las pérdidas en el seguimiento pueden obedecer a la aparición de eventos adversos al medicamento en estudio, a una mala respuesta al mismo, a muerte, etcétera o, al contrario, porque tienen una evolución particularmente benigna y su bienestar hace que no regresen a sus evaluaciones. Así, al ocurrir esto se puede tener una impresión distorsionada de las bondades (o de su ausencia) del tratamiento bajo prueba.

Tal como sucede en la práctica médica cotidiana, los pacientes en los estudios experimentales dejan de tomar sus medicamentos, ya sea por olvido o por efectos indeseables de los mismos. Excluir a estos individuos —que no se adhirieron de manera completa a los lineamientos del protocolo— del análisis puede igualmente generar sesgos. Igual que en el punto anterior, es frecuente que esta falta de apego esté estrechamente vinculada con un pronóstico diferente al de los enfermos cumplidores, tal como se ha llegado a constatar en algunos ensayos clínicos, donde incluso entre los pacientes a quienes se les asignó recibir el placebo tuvieron una mejor evolución los que mostraron un buen apego a su ingesta en comparación con los incumplidos. Al excluir a éstos del análisis uno corre el riesgo de sólo evaluar a pacientes con un mejor pronóstico, alterando la condición de equidad entre grupos dada por la asignación aleatoria de la maniobra.

Es justamente este principio de incluir en el análisis a todos los pacientes, tal como fueron asignados de acuerdo con el sorteo, el que define el análisis de "intención a tratar". Esta estrategia metodológica conserva el valor de la aleatorización, es decir, el de lograr que los pacientes con diferentes pronósticos de la enfermedad queden igualmente distribuidos en los grupos (brazos) del ensayo.

**Evitar que las expectativas de los pacientes y de sus evaluadores influyan en la medición de los desenlaces**

Igual que los investigadores, los enfermos que saben que están recibiendo un nuevo trata-

miento en experimentación generalmente tienen una idea prejuiciada sobre su eficacia. Estas expectativas, optimistas o desfavorables, pueden llegar a distorsionar la medición de los resultados, particularmente cuando su evaluación es mediante indicadores propicios a ser influidos por la subjetividad del informante o del propio evaluador ("datos blandos"). Esto puede ocurrir cuando se miden síntomas (como el dolor), sentimientos de bienestar (como la calidad de vida) o signos clínicos sin definiciones suficientemente objetivas y poco reproducibles. En estas condiciones, y cuando los investigadores conocen quién de sus pacientes del estudio recibe el tratamiento experimental y quién no, se pueden sesgar los resultados al hacer una búsqueda diferencial de los desenlaces o al dar interpretaciones diferentes a los hallazgos.

La mejor manera de evitar este tipo de errores es mediante el enmascaramiento, que no es otra cosa que tratar de que ni el enfermo ni el investigador que lo evalúa (doble cegamiento) sepan si aquél se encuentra recibiendo la terapia en prueba o la intervención de contraste. Esto habitualmente se logra administrando un placebo (o la terapia habitual) con apariencia, sabor y textura indistinguibles de la terapia experimental. De no ser posible esto, y si el tipo de desenlaces a medir lo amerita, habrá que diseñar una estrategia en la que el investigador que los evalúa desconozca a qué grupo de estudio pertenecen sus pacientes.



**Brindar igual atención médica, fuera de la terapia en estudio, a los grupos del ensayo**

Si los cuidados médicos, incluyendo otras terapias diferentes a la que se estudia, se dan de manera diferente en el grupo experimental que en el control se corre el riesgo de comprometer seriamente los resultados del ensayo al cometer el sesgo denominado de "cointervención"; es decir, será difícil distinguir qué tanto de la diferencia (o no diferencia) observada entre los grupos es efecto de la terapia en estudio *vs.* consecuencia de una atención médica diferencial. Una manera de evitar este tipo de error sistemático puede ser mediante el diseño de un ensayo doble ciego.

## ¿QUÉ DETERMINA QUE SUS RESULTADOS SEAN APLICABLES A OTROS GRUPOS DE ENFERMOS?

Lo común es que una de las principales motivaciones que lleva a un investigador a realizar un ensayo clínico es el deseo de que la terapia experimental eventualmente sea utilizada por otros médicos, en otros sitios, como una medida útil y que se brinde para el beneficio real del mayor número de enfermos. Esto dependerá fundamentalmente de tres aspectos:

Primero, de qué tanto se parezcan los enfermos que estudia a los enfermos en quienes posteriormente se quiera extrapolar la experiencia obtenida en el ensayo clínico. Esto depende del rigor en la selección de los sujetos a incluirse en el experimento; mientras más criterios de inclusión sea necesario cumplir, menos posibilidades de que los resultados sean aplicables a otras poblaciones de enfermos. De hecho, el investigador suele enfrentarse al dilema de que por querer estudiar a una población relativamente homogénea de enfermos que le permita incrementar la eficiencia de su estudio (es decir, demostrar con claridad un efecto benéfico de la terapia experimental con el menor número de enfermos y costos posibles) establece criterios de inclusión al estudio demasiado estrictos que lo alejan de las

características del quehacer médico cotidiano: de la "vida real".

Segundo, del desenlace que se haya elegido para valorar la eficacia de la terapia experimental. Habitualmente y por razones de eficiencia, el investigador suele elegir mediciones fisiológicas, bioquímicas, microbiológicas (o de otra estirpe) más fáciles de medir, en corto tiempo y con menos pacientes como sustitutos de desenlaces clínica o socialmente más importantes (como puede ser la calidad de vida o la supervivencia) bajo el supuesto de que aquéllas pueden predecir éstos. Así, no es raro que en los estudios experimentales se evoque el beneficio de una nueva terapia porque ésta mejora las pruebas de función respiratoria, o porque disminuye la carga viral en el plasma, o los niveles séricos de colesterol (por mencionar sólo algunos ejemplos), asumiendo que al lograrlo los enfermos vivirán más tiempo y mejor. Es necesario ser cauteloso en la selección de los resultados a evaluar y que sean realmente importantes para los individuos y para la sociedad; de lo contrario, se corre el riesgo de promulgar la bondad de una terapia nueva porque mejora una medición intermedia, pero que a la larga se demuestre su efecto nocivo en desenlaces clínicamente más importantes o incluso en la supervivencia de los sujetos. Ejemplos de ello es lo sucedido con el uso de hipolipemiantes (como el clofibrato y ciertas estatinas) en la reducción del colesterol sérico o de ciertos antiarrítmicos (como la encainida) posterior a un infarto al miocardio.

Tercero, de que la intervención que se evalúa sea única y precisa, como lo es un medicamento. En este caso, su reproducción por otros investigadores (u otros médicos) es fácil. Por el contrario, hay tratamientos que implican varios elementos cambiantes o que demandan un grado de habilidad muy particular de quien realiza la maniobra (como sería el caso del manejo médico de cierta entidad clínica en un determinado ambiente, una intervención quirúrgica, un programa educativo, etc.), cuyos efectos no necesariamente serán los mismos cuando son realizados por otros médicos o investigadores.



## ¿CÓMO SE ANALIZAN SUS RESULTADOS?

Lo más común es que los ensayos clínicos controlados midan la incidencia de algún evento en los grupos de individuos seguidos en un determinado lapso y que este evento se exprese de manera dicotómica (es decir, la presencia o no del desenlace: infarto al miocardio, recurrencia de una neoplasia, muerte, etc.) como la proporción de sujetos que llegan a presentarlo. Pongamos como ejemplo un estudio en el que 20% (0.20) de los enfermos en el grupo control fallecieron en contraste con sólo 15% (0.15) de los que recibieron el tratamiento en evaluación. En el cuadro II se resume la forma como se puede presentar el efecto de éste, a saber:

1. La diferencia absoluta (o la reducción del riesgo absoluto [RRA]), que se obtiene al sustraer la proporción de individuos que fallecieron en el grupo experimental (Y) de la proporción de individuos que lo hicieron en el grupo control (X): $X - Y = 0.20 - 0.15 = 0.05$ (5%).
2. El riesgo relativo (RR), es decir, el riesgo de morir en los pacientes sometidos a la terapia experimental en relación con

**Cuadro II Mediciones del efecto de un tratamiento en evaluación***

| | | |
|---|---|---|
| Reducción del riesgo absoluto (diferencia de riesgos) (RRA) | X − Y | 0.20 − 0.15 = 0.05 o 5% |
| Riesgo relativo (RR) | Y/X | 0.15/0.20 = 0.75 |
| Reducción del riesgo relativo (RRR) | 1 − (Y/X) × 100<br>o<br>(X − Y)/X × 100 | 1 − 0.75 × 100 = 25%<br><br>0.05/0.20 × 100 = 25% |
| Número de pacientes necesarios a tratar para prevenir un evento (NNT) | 1/(Y − X) | 1/(0.20 − 0.15) = 20 |

* X = riesgo del evento en los pacientes sin el tratamiento (grupo control) 20/100 = 0.20 o 20%.
Y = riesgo del evento en los pacientes con el tratamiento (grupo experimental) 15/100 = 0.15 o 15%.

el de los pacientes en el grupo control: Y/X = 0.15/0.20 = 0.75.

3. El complemento del riesgo relativo (o la reducción del riesgo relativo [RRR]) que se expresa como un porcentaje: [1 − (Y/X)] × 100 = [1 − 0.75] × 100 = 25%. Esta cifra significa que el nuevo tratamiento reduce el riesgo de morir en 25% en relación con lo que ocurre en los pacientes del grupo control; mientras mayor sea la RRR mayor es la eficacia del tratamiento.
4. El número necesario de pacientes a tratar (NNT) indica si el beneficio ofrecido por la nueva terapia retribuye el esfuerzo y costo en su adquisición o implantación.[9] Por ejemplo, una reducción de 25% en el riesgo de morir puede parecer impresionante, pero su impacto en el paciente o en la práctica clínica puede,



**Cuadro III Ejemplo del efecto del pronóstico de la enfermedad en el número necesario de pacientes a tratar**

| | | |
|---|---|---|
| Riesgo de muerte en los pacientes en el grupo control (riesgo basal): X | 1% o 0.01 | 10% o 0.10 |
| Riesgo relativo en los pacientes en el grupo experimental: X/Y | 75% o 0.75 | 75% o 0.75 |
| Reducción del riesgo relativo: 1 − (Y/X) × 100 o (X − Y)/X × 100 | 25% | 25% |
| Riesgo de muerte en los pacientes en el grupo experimental: Y | 0.01 × 0.75 = 0.0075 | 0.10 × 0.75 = 0.075 |
| Reducción del riesgo absoluto: X − Y | 0.01 − 0.0075 = 0.0025 | 0.10 − 0.075 = 0.025 |
| Número de pacientes a tratar para evitar una muerte: 1/(Y − X) | 1/0.0025 = 400 | 1/0.025 = 40 |

sin embargo, ser mínimo. La utilidad de un tratamiento está no sólo en función de la reducción relativa del riesgo sino también del riesgo del desenlace adverso que se quiere prevenir (en nuestro ejemplo, la muerte); de tal forma que mientras menor sea este riesgo mayor será el número necesario de enfermos a tratar con la nueva terapia para prevenir una muerte (es decir, menor su impacto). En el cuadro III se ilustran dos circunstancias: una en la que el riesgo de muerte en una población de enfermos, en un determinado periodo de tiempo, es sólo de 1%; en contraste con otra población con un riesgo mayor (de 10%). En el primer caso, la nueva terapia reduciría el riesgo de fallecer en 25%, es decir, una reducción del riesgo absoluto de 0.0025 (o 25 muertes en 10 000 pacientes tratados). El número necesario de pacientes a tratar se obtiene al calcular la inversa de esta reducción del riesgo absoluto (1/0.0025 = 400); así, sería necesario tratar 400 pacientes durante un tiempo determinado para salvar una sola vida. En cambio, en el segundo caso, una reducción relativa de 25% de muerte en una población en mayor riesgo de morir (de 10%) lleva a una reducción del riesgo absoluto de 0.025 (o 25 muertes en 1 000 pacientes tratados) de tal suerte que se tendría que tratar a sólo 40 individuos para salvar una vida (1/0.025 = 40). Este ejemplo señala un elemento clave en la decisión de implantar una nueva terapia: considerar la magnitud del riesgo del desenlace adverso en los pacientes no tratados con ella. Para una misma reducción del riesgo relativo mientras mayor sea la probabilidad de padecer un evento indeseable si no se trata, mayor será el beneficio con la nueva terapia y menor el número de pacientes que tendremos que tratar para prevenir un evento. Cabe mencionar que en esta decisión debe considerarse también el costo, la factibilidad y el grado de seguridad de la nueva terapia en cuestión. Si, por ejemplo, ésta conlleva un riesgo de 10% de un cierto efecto adverso por el tipo de medicamento, en la población de enfermos con un bajo riesgo de muerte, se tendrían 40 individuos sufriendo el efecto indeseable de la droga por cada vida salvada contra sólo cuatro si se le da a la población de enfermos con un mayor riesgo de muerte. Será, finalmente, el costo de la terapia, así como la gravedad y tipo de consecuencias del efecto adverso del medicamento, lo que nos haga decidir si 4, 40, 400, 4 000 o 40 000 pacientes necesarios a tratar es una cantidad importante o no.

Otro aspecto importante a evaluar en la medición de los resultados de un ensayo clínico es qué tan precisa fue la estimativa del efecto del tratamiento. La verdadera reducción del riesgo es algo que nunca llegaremos a conocer; lo más que podemos alcanzar es llegar a estimarla y el mejor estimado es el valor observado en el estudio (el llamado "estimado puntual"). Mediante el cálculo estadístico del intervalo de confianza (IC) uno puede establecer una zona de valores, alrededor de este estimado puntual, donde pudiera encontrarse el verdadero valor poblacional. Lo habitual, aunque arbitrario, es que se use el IC al 95%; es decir, se establece el intervalo que incluye al verdadero valor de la reducción del riesgo relativo en 95% de las veces. Será raro (con una probabilidad de sólo 5%) que éste se encuentre más allá de los límites de este intervalo; lo que va de acuerdo con lo que convencionalmente se establece como el nivel de significancia estadística (o valor de *p*).

Examinemos un ejemplo. Si en un ensayo clínico se aleatorizaron 100 pacientes al grupo experimental y 100 pacientes al grupo control y se llegan a observar 15 muertes en el primer grupo y 20 en el segundo, el cálculo del estimado puntual de la reducción del riesgo relativo sería de 25%: X = 20/100 o 0.20, Y = 15/100 o 0.15, y [1 − (Y/X)] × 100 = [1 − 0.75] × 100 = 25%. Por lo anteriormente dicho, pudiera ser que el verdadero valor de la RRR fuera significativamente menor o mayor que esta cifra de 25%, la que se obtuvo de una diferencia de tan sólo cinco muertes entre los dos grupos, y hasta pensarse que el tratamiento no fuera eficaz (una RRR de 0%) e incluso perjudicial (una RRR con un valor negativo). De hecho así es: estos resultados son consistentes tanto con una RRR de menos 38% (es decir, que a los pacientes que reciben el nuevo tratamiento tuvieran un riesgo de morir 38% mayor que los pacientes en el grupo control) y una RRR de casi 59% (es decir, que los pacientes en el grupo experimental tuvieran un riesgo menor de morir de casi 60%). En otras palabras, el intervalo de confianza al 95% (IC 95%) de este estimado de la RRR es de menos 38% a 59%, y el estudio en realidad no nos ayuda a decidir si el nuevo tratamiento es útil. Ahora, supongamos que en lugar de 100, se hubieran sorteado 1 000 pacientes por grupo y que se hubiera observado la misma proporción de desenlaces, es decir, 150 muertes en el grupo experimental (Y = 150/1 000 = 0.15) y 200 muertes en el grupo control (X = 200/1 000 = 0.20). Nuevamente, el estimado puntual de la RRR sería 25%: [1 − (Y/X)] × 100 = [1 − (0.15/0.20)] × 100 = 25%. Se puede predecir que, en este ensayo con un número bastante mayor de individuos estudiados, el verdadero valor de la RRR esté más cerca del valor de 25%; efectivamente, el IC 95% de la RRR en estos datos va de 9 a 41%.

Lo que estos ejemplos nos muestran es que mientras mayor sea el número de participantes en el ensayo clínico mayor será el número de eventos observados y mayor la certidumbre de que el valor verdadero de la RRR (o de cualquier otra medida de eficacia) está cercano al valor obtenido en el estudio. En el segundo ejemplo anterior, el valor posible más bajo de la RRR fue de 9% y el más alto, de 41%. El estimado puntual (en este caso 25%) es probablemente el que más se acerca al valor real (poblacional) de la RRR. Mientras más lejanos los valores del estimado puntual menos probable es que sean consistentes con el valor observado y aquéllos más allá de los límites del IC son valores con muy pocas posibilidades de representar la verdadera RRR, dado el estimado puntual (la RRR observada).

Siendo que mientras más grande es el tamaño de la muestra estudiada más estrecho es el IC, ¿cuál es un número suficiente de sujetos a estudiar? Cuando en un ensayo clínico se concluye que los resultados fueron positivos (es decir, que muestra que el tratamiento fue eficaz) es porque el valor inferior del IC del estimado de la RRR (el más bajo y aún consistente con los resultados del estudio) es "clínicamente importante", es decir, lo suficientemente grande como para que el tratamiento sea prescrito a los enfermos. En esta circunstancia puede decirse que el tamaño de la muestra estudiada fue suficiente. Si, en cambio, se considera que el límite inferior de este IC no es "clínicamente importante", entonces el estudio no puede ser considerado como definitivo, a pesar de que la diferencia entre tratamientos sea estadísticamente significativa (es decir, que excluya que la RRR sea de cero).

El IC también ayuda a interpretar los estudios con resultados "negativos", en los que los autores concluyen que el tratamiento experimental no mostró ser mejor que el del grupo control: si el límite superior del IC muestra que la RRR pudo haber sido "clínicamente significativa" quiere decir entonces que el estudio no logró excluir un efecto importante de la nueva

terapia. En el primer ejemplo anterior, el límite superior del IC fue una RRR de 59%; de tal forma que el beneficio del tratamiento pudo haber sido sustancial y se concluiría que aunque los investigadores no lograron demostrar que la terapia experimental fuera mejor que el placebo tampoco pudieron descartar un efecto importante de ella. Esto es lo que se conoce como un estudio con poca precisión (poder o sensibilidad) por haber estudiado un tamaño de muestra insuficiente.

Cabe señalar la asociación entre el IC y el valor de significancia estadística, o valor de *p*. Si éste es igual o mayor a 0.05 (5%) quiere decir que el límite inferior del IC 95% de la RRR es el valor nulo (cero), o un valor negativo, y el RR, de uno, o menor; es decir, que no se logra descartar la hipótesis nula de no diferencia entre el tratamiento experimental y el placebo (o terapia en el grupo control). Conforme el valor de *p* disminuye (menor a 5%) el límite inferior del IC 95% de la RRR es mayor a cero y se dice que la diferencia entre tratamientos mostró ser estadísticamente significativa; es decir que, si la hipótesis nula es cierta, es muy pequeña la posibilidad de ver una diferencia de esta magnitud o mayor. La prueba estadística para el cálculo del valor de *p* dependerá del tipo de variable, de cómo esté expresado el evento desenlace (dicotómica, ordinal, continua) y su descripción no es motivo de este capítulo.

## ¿ES ÉTICO HACERLOS?

Realizar un ensayo clínico controlado puede generar inquietudes de orden ético. Algunos pacientes se pueden llegar a incomodar por el hecho de ser interrogados o examinados para propósitos diferentes a los de su estricta atención médica, o bien, al saber que investigadores que no son sus médicos llegan a tener acceso a la privacía de su expediente clínico. El diseño experimental significa que los tratamientos son asignados por un investigador, no por su médico tratante ni elegidos por el propio paciente. Cuando la asignación es mediante un sorteo o cuando hay un doble enmascaramiento (es decir, que ni el paciente ni sus médicos que lo atienden saben a qué medicamento se está sometiendo) los aspectos éticos se vuelven más complejos e incluso más importantes que los científicos.

Para que un estudio clínico experimental sea éticamente justificable debe cumplir la premisa de que, al momento del inicio del estudio, no haya evidencia de que alguno de los tratamientos ofrecidos en cada brazo (o grupo del ensayo, incluyendo el control) sea superior al(os) otro(s). No es raro que al plantearse la conducción de un estudio clínico experimental surja la inquietud de que a los participantes no se les esté ofreciendo la mejor opción terapéutica conocida; es decir, que se les está privando de una terapia experimental novedosa y superior a lo ya conocido o, por el contrario, que se les esté exponiendo a un nuevo tratamiento de dudosa utilidad y seguridad. En efecto, el estudio no sería ético si se tiene la suficiente información de que uno de los tratamientos es más eficaz (o dañino); pero si en verdad se desconoce si el tratamiento en evaluación es el óptimo, no habría justificación para dicha preocupación; de hecho, podría argumentarse la falta de ética al estar prescribiendo abierta y rutinariamente un tratamiento cuya utilidad (y seguridad) relativa se desconocen porque se adoptó de manera prematura, antes de su evaluación formal. En este punto puede existir la dificultad de definir a partir de cuándo la evidencia es "suficiente" o "convincente" para justificar o no la realización de un estudio experimental en humanos; desde lue-

go que la perspectiva de un clínico (que procura el máximo beneficio de un individuo en particular) será diferente que la del investigador clínico (quien busca la "verdad" con rigor científico). Por otra parte, es habitual que en el proceso de desarrollo de un nuevo medicamento se le exija a quien lo manufactura que presente información de su eficacia obtenida en pequeños estudios no controlados (la llamada fase II) antes de encaminarse a un ensayo clínico controlado (fase III).

Aunque estos dilemas éticos pueden llegar a ser de difícil solución, en muchos centros de investigación del mundo el consentimiento informado y los comités institucionales de investigación en humanos han venido a constituir importantes salvaguardas de la ética. El consentimiento informado es requerido por la mayoría de las agencias financiadoras de proyectos de investigación experimental en seres humanos así como por las revistas científicas y habitualmente es obligatorio que sea firmado por el participante. El consentimiento informado implica que al individuo se le da la oportunidad de preguntar y enterarse de la naturaleza y detalles del estudio y se hace de su conocimiento que goza de entera libertad para decidir si participa en el estudio clínico así como de que no habrá ningún tipo de represalia (o cambio en su atención médica) si decide no hacerlo, o si abandona el estudio antes de lo programado. Sin embargo, a pesar de estos requisitos, y dada la creciente complejidad de los protocolos de los estudios experimentales, no es rara la circunstancia en la que el sujeto participante no llegue a estar completa y claramente informado. Los comités institucionales de investigación en humanos son cuerpos colegiados que suelen estar constituidos por investigadores, médicos clínicos, enfermeras, administradores, abogados y representantes de la comunidad y su tarea consiste en revisar escrupulosamente el protocolo de investigación, constatando que el planteamiento de la pregunta de investigación y el diseño sean adecuados (después de todo no es ético realizar un estudio con importantes errores metodológicos) y asegurando la preservación de la integridad y seguridad de los participantes en el estudio y que la carta de consentimiento informado esté correctamente elaborada. El comité puede también solicitar que los datos del estudio sean analizados periódicamente por un comité externo y ante la eventualidad de encontrar una clara diferencia en la eficacia entre los tratamientos, que se interrumpa anticipadamente el estudio y se publiquen los hallazgos.

Finalmente, cabe comentar sobre la justificación de cuándo dar un placebo a los enfermos en el grupo control. Esto ocurre en dos circunstancias: una, cuando no existe tratamiento alguno que convincentemente haya mostrado ser eficaz en modificar favorablemente el curso de la enfermedad; es decir, que en una buena práctica médica en condiciones habituales un médico no prescribiría medicamento específico alguno (tal es el caso de la evaluación del zanamivir, un inhibidor de la neuraminidasa, en el tratamiento de la influenza o del uso de la warfarina en la prevención de fenómenos embólicos en la fibrilación auricular sin valvulopatía). La segunda circunstancia es cuando sí existe un tratamiento estándar efectivo pero se desea evaluar si la adición de otro medicamento ofrece una ventaja adicional. En este caso a ambos grupos de enfermos se les administra la terapia estándar, pero al experimental se le adiciona el nuevo medicamento en estudio y, en cambio, al grupo control un placebo indistinguible del nuevo medicamento (tal es el caso de la evaluación de si los ácidos grasos 3-omega, como tratamiento complementario, mejoran la funcionalidad de los enfermos con artritis reumatoide activa que además reciben medicación antinflamatoria).

En cambio, hay circunstancias en las que sería una grave falta ética dar únicamente un placebo y privar de un tratamiento activo ya

ampliamente establecido y aceptado como útil a los individuos en el grupo control. Difícilmente alguien que desea probar que una cierta terapia es eficaz para mejorar el pronóstico de los pacientes sintomáticos por infección por el VIH la compararía con un grupo de enfermos en quienes se omitiera el tratamiento antirretroviral estándar en la actualidad: dos análogos nucleósidos inhibidores de la transcriptasa reversa más un inhibidor de proteinasas.

## ¿CUÁLES SON SUS VENTAJAS Y DESVENTAJAS?

Respecto a los estudios observacionales de cohortes, la principal ventaja que ofrecen los ensayos clínicos controlados (especialmente los que son mediante un sorteo) es que asignan la maniobra experimental (el tratamiento en prueba) independientemente de determinantes pronósticos o de la selección de la muestra; de tal suerte que disminuye el riesgo de sesgos de susceptibilidad diferencial (confusores) o de un muestreo distorsionado. Además, el diseño experimental facilita el enmascaramiento de los individuos participantes y de sus evaluadores (doble cegamiento), ya que en un estudio observacional los tratamientos generalmente son decididos y prescritos abiertamente por los médicos tratantes.

Por otra parte, la posibilidad real de realizar los estudios experimentales puede verse seriamente comprometida por razones de índole práctica (logística) de costos y de ética. Además, en un ensayo clínico se puede llegar a una situación tal de control de variables (homegeneidad de la muestra estudiada, regulación del apego al tratamiento, atención y seguimiento del enfermo, etc.) que haga que sus resultados se alejen de la realidad del quehacer clínico cotidiano, restringiendo seriamente su extrapolación al resto de los enfermos; es decir, que en la compulsión por diseñar estudios metodológicamente impecables no se logre responder a las preguntas originales.

A pesar de esto último, los ensayos clínicos aleatorizados permanecen como el estándar de oro (o paradigma) en la evaluación de la utilidad de nuevos tratamientos. A pesar de no ser infalibles y de, en ocasiones, dejar resquicios de incertidumbre, los estudios experimentales han contribuido enormemente a la evolución hacia terapias y medidas profilácticas más eficaces y seguras.

Además, el entendimiento de los principios científicos que rigen a los ensayos clínicos controlados le ha ayudado al médico a adquirir la destreza para discriminar con rapidez los estudios clínicos cuyas conclusiones son poco sustentables, por ser estudios de mal diseño y realización, de aquellos realmente válidos y aplicables a su propia práctica. En la medida en que los médicos sólo demos credibilidad a los trabajos publicados con solidez científica (es decir, que nos transformemos de lectores pasivos y puramente receptivos en evaluadores críticos) elevaremos la calidad de nuestras decisiones y el grado de beneficio que reciban nuestros enfermos.



## REFERENCIAS

1. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. New York, NY: John Wiley & Sons, INC, 1982, p 40-44.

2. Stevens DA, Jordan GW, Waddell TF, Merigan TC. Adverse effect of cytosine arabinoside on disseminated zoster in a controlled trial. N Engl J Med. 1973;289:873-878.
3. Crosby NJ, Deane KH, Clarke CE. Amantadine for dyskinesia in Parkinson's disease. Cochrane Database Syst Rev. 2003;(2):CD03467.
4. Leather RA, Sullivan SN. Iced gastric lavage: a tradition without foundation. CMAJ. 1987;136:1245-1247.
5. Cronin L, Cook DJ, Carlet J, Heyland DK, King D, Lansang MA, *et al*. Corticosteroid treatment for sepsis: a critical appraisal and meta-analysis of the literature. Crit Care Med 1995;23:1430-1439.
6. Tindall R. A Closer Look at plaspapheresis in multiple sclerosis: the cons. Neurology 1988;38(suppl 2):53-56.
7. O'Nelly P. Acute otitis media. BMJ 1999;319:833-5.
8. Haynes RB, Mukherjee J, Sackett DL, Taylor DW, Barnett HJ, Peerless SJ. Functional status changes following medical or surgical treatment for cerebral ischemia. Results of the extracranial-intracranial bypass study. JAMA. 1987;257:2043-2046.
9. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995;310:452-454.

# V Ensayos clínicos aleatorizados

*Eduardo Lazcano Ponce*
*Eduardo Salazar Martínez*
*Pedro Gutiérrez Castrellón*
*Angélica Ángeles Llerenas*
*Adolfo Hernández Garduño*
*José Luis Viramontes*

*Me levanté muy temprano para visitarlos, con la esperanza de encontrar a aquellos a quienes había administrado un medicamento digestivo. Sintiendo un poco de dolor, sus heridas no habían crecido o inflamado y habían podido dormir durante la noche. Los otros a quienes yo había aplicado un aceite hirviendo tuvieron fiebre con mucho dolor y protuberancias alrededor de sus heridas. Entonces yo determiné nunca otra vez quemar así tan cruelmente a los pobres heridos por arquebús.*[1]

Ambroise Paré (1510-1590).



## INTRODUCCIÓN

Como se ha señalado en capítulos previos, una de las formas más comunes de clasificar los estudios epidemiológicos consiste en agruparlos, según se asigne la exposición, en experimentales y observacionales. La primera característica de los estudios experimentales es que el investigador manipula o asigna la exposición en forma intencional y aleatoria. En cuanto a su característica de temporalidad, los estudios experimentales son de carácter prospectivo y, por el número de observaciones sucesivas realizadas durante el periodo de estudio, se catalogan como longitudinales prospectivos. Asimismo, y a diferencia de los diseños observacionales —donde los criterios de selección de la población se basan en la presencia del evento de estudio (casos y controles) o exposición (cohorte)—, los experimentales persiguen homogeneizar las poblaciones de estudio a través del proceso de aleatorización, de forma que puedan compararse en cuanto a su condición de enfermedad o las características biológicas y sociodemográficas de los sujetos, y las unidades de análisis pueden ser individuos o grupos (como las de tipo *cluster* o intervenciones comunitarias).

El diseño experimental clásico puede definirse a partir de varias características.[2] La primera tiene que ver con el control de las condiciones bajo estudio, que básicamente consisten en la selección de los sujetos; la manera como se administra el tratamiento; la forma en que se llevan a cabo las observaciones; los instrumentos que se utilizan para realizar las mediciones, y los criterios de interpretación. Todas ellas deben implementarse de la manera más uniforme y homogénea posible. La segunda

es que debe haber una maniobra de intervención bajo estudio y, por lo menos, un grupo control. La tercera, que los participantes en el estudio deben asignarse aleatoriamente a los grupos de intervención; esto es, que ningún investigador, médico participante o sujeto de estudio, debe influir, directa o indirectamente, en la toma de decisiones respecto del tratamiento que recibirán los pacientes. La cuarta indica que, si las variables de estudio son confusores conocidos o existen variables pronósticas relevantes (edad, sexo, grupo étnico o severidad de la condición clínica estudiada), la población de sujetos de estudio debe estratificarse en subgrupos, con el fin de garantizar la distribución simultánea de manera equilibrada entre los grupos de comparación o tratamiento. La quinta y última característica es que en un diseño experimental se requiere que el evento de interés (*outcome*) sea perfectamente definido y cuantificado antes y después de haber realizado la intervención.[3] Es necesario considerar que el vocablo ensayo es equiparable, en este tipo de estudios, a la palabra *experimento*.

## DEFINICIÓN DE ENSAYO CLÍNICO CONTROLADO ALEATORIZADO (ECCA)

Un ensayo clínico es un experimento controlado en voluntarios humanos que se utiliza para evaluar la seguridad y eficacia de tratamientos y/o intervenciones contra enfermedades y problemas de salud de cualquier tipo; así como para determinar efectos farmacológicos, farmacocinéticos y/o farmacodinámicos de nuevos productos terapéuticos, incluyendo el estudio de sus reacciones adversas. Es decir, un ensayo clínico es un experimento con pacientes como sujetos de estudio, dentro del cual, cuando se prueba un nuevo medicamento, se comparan por lo menos dos regímenes de tratamiento, uno de los cuales se utiliza como control. Esta definición aplica también a los experimentos controlados para probar vacunas con fines preventivos en población sana al inicio del ensayo. Un ejemplo lo constituye el estudio realizado por Oxman[4] en 38 546 adultos sanos mayores de 60 años, quien aplicó una vacuna de virus atenuados de herpes zoster y, posterior a un seguimiento mayor de tres años de los grupos de intervención y placebo, observó una reducción en la incidencia de la infección de 51%. En este contexto, se distinguen dos tipos de controles: los pasivos y los activos. Un control pasivo utiliza placebo en un ensayo de agentes terapéuticos, lo que significa la inclusión de un producto inocuo, cuya preparación por sí misma es similar en todas las características (presentación, tamaño, color, textura y sabor) al otro agente terapéutico, con la única diferencia de que no contiene el principio activo. En algunos casos, en los que desee demostrarse que la preparación es equivalente o superior al producto estándar existente, y para proteger a pacientes que necesitan medicación por prescripción médica, deberá emplearse un control activo.[5]

Los ensayos clínicos controlados aleatorizados (ECCA) se consideran el paradigma de la investigación epidemiológica, porque son los diseños que más se acercan a un experimento por el control de las condiciones bajo estudio, y porque pueden establecer relaciones causa-efecto, si las siguientes estrategias se establecen eficientemente: a) asignación de la maniobra de intervención mediante mecanismos de aleatorización en sujetos con características homogéneas, que permite garantizar la comparabilidad de poblaciones; b) utilización de un grupo control que permite la comparación no sesgada de

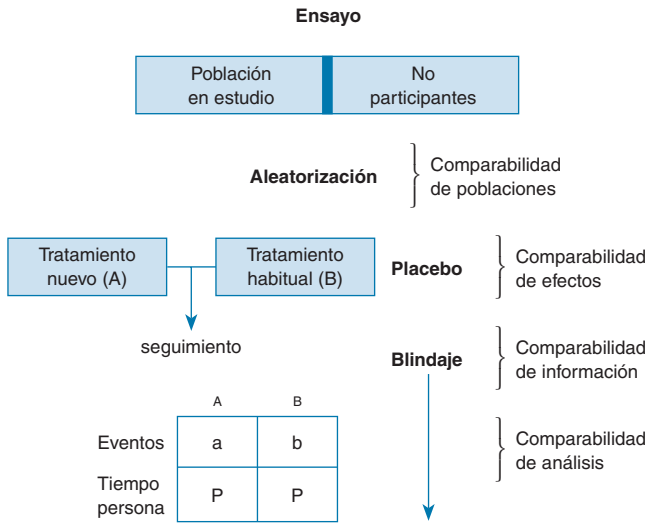



**Figura 1** **Características básicas de un ensayo clínico aleatorizado**

efectos de dos posibles tratamientos: el nuevo y el habitual (o bien, el placebo); c) blindaje de los grupos de tratamiento, que permite minimizar los posibles sesgos de información y da lugar a la comparabilidad de información, y d) incorporación de las estrategias descritas previamente, lo que permite la comparabilidad de mediciones en el análisis. Estas estrategias se aplican siempre y cuando se cumpla el principio de intercambiabilidad, esto es, que si se invierte y asigna la maniobra al grupo control, los resultados deben ser los mismos (figura 1).

## CLASIFICACIÓN DE LOS ENSAYOS CLÍNICOS

Los ensayos clínicos se plantean en forma muy diversa, por lo que es necesario establecer criterios de clasificación. Cuando se estudian por la estructura de tratamiento pueden agruparse en diseños paralelos, de tratamiento sucesivo y ensayos alternativos. En relación con el enfoque de enfermedad, adicional a los ensayos de tratamientos terapéuticos, pueden ponerse en práctica ensayos de prevención primaria y secundaria. En cuanto al enfoque de tratamiento, los ECCA estudian efectos de nuevos medicamentos, nuevas alternativas quirúrgicas, suplementación nutricional, entre otros tipos de intervención. Asimismo, por el tipo de aleatorización, pueden ser aleatorizados y no aleatorizados. Cuando menos, existen tres tipos de asignación de la intervención: fija, dinámica y adaptativa. Por el tamaño de muestra, los ECCA pueden clasificarse en fijos y secuenciales; finalmente, por el número de sedes, pueden ser de sitio único y multicéntricos (cuadro I). A continuación se describirán los ECCA por estructura de tratamiento.

**Cuadro I Clasificación de los ensayos clínicos controlados**

| *Criterios de clasificación* | *Tipo de ensayo clínico* |
|---|---|
| Estructura del tratamiento | a) Diseño paralelo<br>b) Diseño del tratamiento sucesivo<br>• Variantes<br>• Diseño del tratamiento de reemplazo<br>• Diseño cruzado *(crossover)*<br>c) Diseño de ensayos alternativos<br>• Diseño factorial<br>• Ensayos de equivalencia<br>• Aleatorización por conglomerados *(cluster)* |
| Enfoque de la enfermedad | a) Ensayo de tratamiento<br>b) Prevención primaria<br>c) Prevención secundaria |
| Enfoque del tratamiento | a) Ensayo de drogas<br>b) Cirugía<br>c) Dieta<br>d) Otros |
| Tipo de aleatorización | a) Aleatorizados<br>b) No aleatorizados |
| Tipo de asignación | a) Fija<br>b) Dinámica<br>c) Adaptativa |
| Por el tamaño de muestra | a) Fijo<br>b) Secuencial |
| Por el número de sedes | a) Centro único<br>b) Multicéntrico |



## CLASIFICACIÓN SEGÚN LA ESTRUCTURA DEL TRATAMIENTO

*ECCA con diseño paralelo:* en los ECCA de tipo paralelo, los sujetos de estudio siguen el tratamiento, al que han sido asignados aleatoriamente, durante el tiempo que dure el ensayo.

*ECCA con diseño de tratamiento sucesivo:* en los ECCA de tratamiento sucesivo cada sujeto es asignado al azar a un grupo que sigue una secuencia de tratamiento previamente determinada, de manera que cada persona recibe más de un tratamiento. La forma más frecuente es el diseño de tratamiento sucesivo en dos periodos, con un primer tratamiento seguido de un segundo. Entre el primer tratamiento y el segundo, se deja un periodo sin tratamiento, de manera que se disipen los efectos residuales del primero. A este respecto, existen

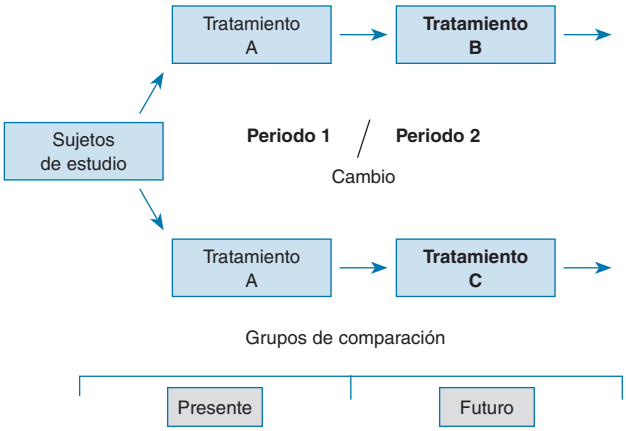



**Figura 2** **Diseño de tratamiento de reemplazo**

básicamente dos tipos de ECCA de diseño de tratamiento sucesivo: el de tratamiento de reemplazo y el cruzado.

*Diseño de tratamiento de reemplazo:* se utiliza para recolectar datos sobre los efectos de cambiar un tratamiento A por uno de dos tratamientos alternativos, por ejemplo, tratamiento B o tratamiento C. Los sujetos de estudio se dividen en dos grupos iguales. Ambos reciben el tratamiento A durante un primer periodo. Las observaciones que se efectúan entre los pacientes tratados con A y B se comparan con los resultados observados entre los pacientes tratados con A y C (figura 2).

*Diseño cruzado:* en este caso, el grupo 1 recibe el tratamiento A durante un primer periodo y el tratamiento B en el segundo. El grupo 2 recibe los tratamientos en orden inverso al grupo 1. Los diseños cruzados permiten ajustar las variaciones de persona a persona, haciendo que cada sujeto sirva como su propio control. En este diseño, se exige con frecuencia un menor número de sujetos en relación con otros (figura 3). Un ejemplo de esta estrategia es el estudio realizado en México para evaluar el impacto nutricional de una intervención comunitaria sobre el crecimiento en niños de bajo ingreso económico, menores de 12 meses de edad, en seis estados del centro del país.[6] Mediante un proceso de aleatorización, 205 comunidades fueron asignadas para el grupo de intervención y 142 comunidades lo fueron para el grupo de intervención cruzada. El grupo de intervención recibió el programa durante los dos primeros años de vida (primer y segundo periodos), mientras que el grupo de intervención cruzada solamente lo recibió durante el segundo año de vida (segundo periodo). Los resultados mostraron un crecimiento mayor en el grupo de intervención (26.4 cm) que en el grupo de intervención cruzada (25.3 cm), así como valores medios de hemoglobina mayores (11.12 *vs.* 10.9, respectivamente).

## Clasificación de diseños alternativos

*Diseño factorial:* la evaluación de dos o más intervenciones en el mismo ECCA puede ponerse en práctica por medio de un diseño de tipo paralelo. Sin embargo, es necesario au-

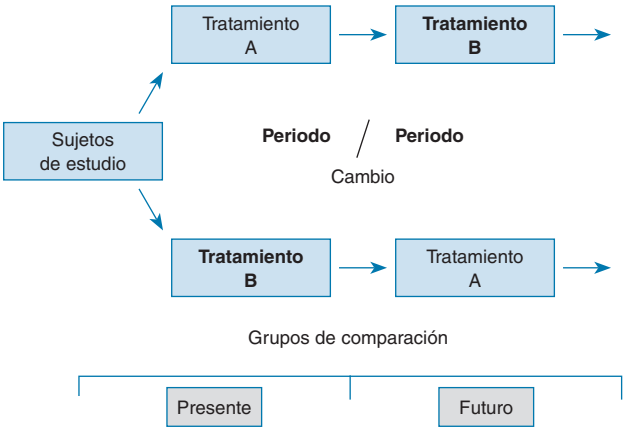


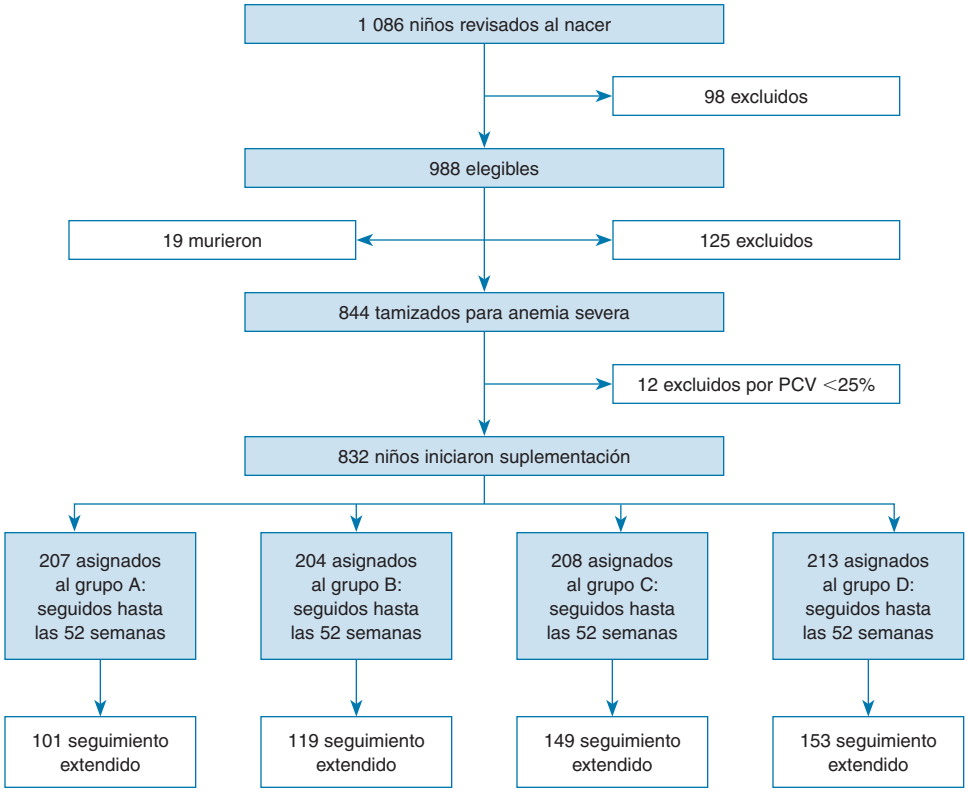

**Figura 3** **Diseño de tratamiento cruzado (*crossover*). Cada sujeto sirve como su propio control**



mentar el tamaño de la muestra, lo cual puede ser ineficaz, especialmente, si también existe el interés por considerar combinaciones de los tratamientos. El diseño alternativo en esta situación es de tipo factorial, en el que pueden asignarse de manera aleatoria dos o más intervenciones en forma independiente, siempre y cuando no exista una interacción; de tal manera que los sujetos pueden, o bien no recibir ninguna intervención, o una de ellas o eventualmente todas.[7] En la figura 4 se refiere un ejemplo de un ensayo factorial, donde cada paciente es aleatorizado dos veces para recibir dos tratamientos en el mismo ensayo;* esto es, bajo el supuesto de que no hay interacción, dos experimentos pueden ser conducidos en uno. Se trata de la evaluación de una intervención para prevenir malaria y anemia, con quimioprofilaxis de deltaprim y hierro, respectivamente. Los autores establecieron con estos resultados que la quimioprofilaxis de malaria durante el primer año de vida es efectiva en la prevención de malaria y anemia, y que la suplementación con hierro es efectiva para prevenir anemia severa, sin incrementar la susceptibilidad a malaria.[8]

*Diseño de equivalencia:* se pone en práctica para demostrar que dos tratamientos son efectivamente similares respecto a la respuesta del paciente. Son diseños no sesgados que evalúan diferencias en tratamiento cercanas a cero y con un estrecho intervalo de confianza. Se ponen en práctica, porque existen tratamientos que pueden diferir en seguridad, efectos adversos, conveniencia de administración, costos, entre otras características. La "equivalencia" tiene importancia para el uso subsiguiente de uno o ambos tratamientos. Un ejemplo de ello es el ensayo clínico realizado en Japón para probar la eficacia del maleato de timolol con sorbato de potasio (MTSP) comparado con maleato de timolol (MT) en sujetos con hipertensión

* El diseño factorial, en este caso, hace la doble aleatorización en un solo tiempo de manera tal que, cuando se forman los cuatro grupos, la doble aleatorización ya se llevó a cabo. Así se garantiza que los sujetos estén potenciados por el doble proceso. Pero el procedimiento se efectúa en un solo tiempo antes de la formación de los grupos.




| Grupo | Número de episodios | Incidencia por años persona |
|---|---|---|
| **Anemia Severa** | | |
| A: Placebo | 81 | 0.62 |
| B: Hierro (2mgs/kg/día) + placebo | 58 | 0.42 |
| C: Placebo + deltaprim | 39 | 0.25 |
| D: Hierro (2mgs/kg/día) + deltaprim | 31 | 0.20 |
| **Episodio clínico de malaria** | | |
| Placebo | 81 | 0.71 |
| Hierro (2mgs/kg/día) + placebo | 75 | 0.63 |
| Placebo + deltaprim | 42 | 0.29 |
| Hierro (2mgs/kg/día) + deltaprim | 36 | 0.24 |

**Figura 4** **Una intervención para prevenir malaria y anemia. Un ejemplo de ensayo clínico factorial**

ocular, incluyendo aquellos con glaucoma de ángulo abierto, con edad igual o mayor a 18 años, y afectación ocular uni o bilateral (presión ocular igual o mayor a 22 mm de Hg). En un grupo se administró MT a nivel ocular, una vez al día, y en el otro, MTSP dos veces al día, durante 12 meses. Al final del periodo del estudio, 95% de los pacientes recibieron los medicamentos asignados. Los resultados fueron similares entre ambos grupos: se encontró una reducción de la presión ocular en ambos brazos del estudio.[9]

## FASES DE UN ECCA PARA EVALUAR EL EFECTO DE NUEVOS FÁRMACOS

La investigación clínica de evaluación de un nuevo agente terapéutico (incluidas las vacunas) previamente no evaluado, generalmente se divide en cuatro fases (cuadro II). Aunque las fases pueden ser conducidas secuencialmente, en algunas situaciones pueden traslaparse.

La fase I incluye el inicio de estudio de un nuevo agente farmacológico en un grupo de 20-80 sujetos, los cuales se vigilan de cerca y pueden clasificarse como sujetos sanos o enfermos. Esta fase del estudio se diseña para determinar las acciones farmacológicas, el metabolismo de las drogas en humanos, así como los mecanismos de acción, las reacciones adversas asociadas con el incremento de dosis, y, si es posible, la evidencia temprana de su



**Cuadro II** **Fases de un ensayo clínico controlado aleatorizado para evaluar efectos terapéuticos de nuevos fármacos**

| | *F A S E S* | |
|---|---|---|
| | *Fases de un ensayo clínico aleatorizado* | *Ejemplo de Caso La prevención de infección persistente por Virus de Papiloma Humano* |
| *Estudios preclínicos* | Estudios experimentales preliminares del efecto de nuevos fármacos evaluados en animales, particularmente ratones y conejos, entre otros. | Estudios en modelos animales, desarrollados en la década de los 90s, mostraron que la inmunización sistémica con vacuna con VLP (partículas parecidas a virus de papiloma humano) compuesta de L1, que constituye el principal componente de una proteína viral estructural, es capaz de conferir protección en contra de cambios experimentales producidos por los virus de papiloma humano homólogos.[10] |
| *Fase I* | Son diseñados para establecer los efectos de una nueva droga en humanos. Estos estudios son habitualmente conducidos en pequeños grupos de sujetos saludables, especialmente para determinar posibles efectos tóxicos, absorción, distribución y metabolismo. | Cuando una vacuna monovalente L1 de VLP contra VPH 16 fue inicialmente probada en humanos, se observó buena tolerabilidad y alta inmunogenicidad. La respuesta sérica de títulos de anticuerpos fue inicialmente 40 veces más alta comparada con una infección natural. |

**Cuadro II** ***Continuación***

| | *F A S E S* | |
|---|---|---|
| | *Fases de un ensayo clínico aleatorizado* | *Ejemplo de Caso La prevención de infección persistente por Virus de Papiloma Humano* |
| *Fase II* | Después de haber completado exitosamente la fase I, se evalúa la seguridad y eficacia en una población mayor de individuos que están relacionados con la enfermedad o condición. | Posterior a estos ensayos incipientes, se desarrolló un ensayo clínico doble ciego, controlado por placebo de fase II, para evaluar la eficacia de una vacuna profiláctica cuadrivalente L1 de VLP contra los tipos 6, 11, 16 y 18, en mujeres jóvenes. 277 mujeres recibieron la vacuna y a 275 de ellas se les aplicó el placebo. La eficacia de la vacuna para proteger contra la incidencia y persistencia de la infección, así como de los cambios clínicos producidos por la misma, fue mayor de 90%.[11] |
| *Fase III* | La tercera y última fase de pre-aprobación en la evaluación de una droga se realiza en un grupo más grande de individuos relacionados con la enfermedad. La fase II usualmente prueba la nueva droga, en comparación con la terapia estándar que normalmente se usa para el evento de estudio. | Reportes preliminares de estudios multicéntricos de la fase III de ensayos clínicos de eficacia de una vacuna profiláctica contra VPH, en más de 35 000 mujeres de 40 áreas geográficas a nivel mundial, sugieren fuertemente una protección elevada contra la persistencia de la infección por VPH y neoplasia intraepitelial cervical. Sin embargo, la duración de la protección de esta vacuna no es conocida aún, la respuesta de anticuerpos inducida es probablemente VPH por tipo específica y la inmunización deberá ocurrir antes de la exposición al VPH. Una segunda generación de vacunas incluirá un antígeno temprano para protección postexposición.[12] |
| *Fase IV* | Después de que una droga ha sido aprobada, la fase IV es conducida para comparar la droga en relación con otros productos existentes, explorar sus efectos en pacientes de la población general, o para cuantificar adicionalmente la presencia de otros posibles eventos adversos. | La solicitud de aprobación del uso de la vacuna profiláctica contra Virus de Papiloma Humano está en curso y la disponibilidad comercial de este producto será inminente. |

efectividad. También incluye estudios en que los nuevos agentes farmacológicos son utilizados como herramientas de investigación para explorar fenómenos biológicos o el proceso de enfermedad. Durante esta fase, se cuantifican ampliamente los efectos farmacocinéticos y farmacológicos que permitirán planear la fase subsiguiente.

La fase II incluye los estudios clínicos controlados, los cuales tienen como finalidad evaluar la efectividad de las drogas en ciertos pacientes con la enfermedad o condición bajo

estudio; determinar los efectos adversos más comunes, y los riesgos asociados con el uso de estos nuevos agentes farmacológicos. Esta fase debe estar bien controlada, cercanamente vigilada y conducida en un pequeño número de sujetos. Esta fase puede subdividirse en fase II-A, donde se decide si el tratamiento u otro procedimiento en particular son suficientemente efectivos para justificar un estudio adicional. Para ello, se fija un nivel de efectividad, a partir del cual se evalúa la posibilidad de encontrar 95% de éxitos; en otras palabras, sólo se admite 5% de fracasos. La fase II-B se desarrolla con el propósito de estimar la efectividad y magnitud de la misma. Con esta información es posible planear tamaños de muestra en estudios de fase III.

La fase III se realiza cuando existe evidencia preliminar que sugiere efectividad del nuevo agente farmacológico que se ha obtenido, y se pretende ganar información adicional acerca de la seguridad y efectividad necesaria para evaluar la relación beneficio-riesgo. Esta fase de estudio se desarrolla generalmente con un gran número de sujetos. Los estudios clínicos fase IV incluyen todas las investigaciones realizadas después de la aprobación del medicamento; en otras palabras, son los estudios de medicamentos en uso rutinario, también conocidos como estudios de postmercadeo.[13,14] Su objetivo está muy definido, y consiste en obtener conocimiento adicional de la eficacia y seguridad de un medicamento.[13] La información obtenida acerca de un medicamento en los estudios fase I-III, no proporciona bases suficientes para establecer conclusiones finales acerca del valor clínico de un medicamento posterior a su comercialización. En comparación con la fase III, la cual tiene un tipo de diseño de estudio del ECCA, la fase IV requiere de diferentes diseños como reportes de casos, series de casos, estudios de observación comprensiva, estudios de casos y controles, estudios de cohorte, análisis de perfil de prescripción y de reporte de eventos adversos, análisis comparativo de bases de datos y estudios de costo beneficio.[14]



## CARACTERÍSTICAS METODOLÓGICAS DE LOS ECCA

Al definir la hipótesis primaria de un ECCA, se ha sugerido que los investigadores no sólo establezcan una hipótesis nula de la falta de efectos del tratamiento en la comparación de grupos, sino que, con base en una revisión sistemática, puedan elegir algunas hipótesis secundarias alternativas, claramente definidas antes de iniciar el estudio. Asimismo, para evaluar si el diseño tiene la capacidad de responder a la pregunta de investigación, deben considerarse los siguientes aspectos: a) definición del evento resultado primario; b) disponibilidad del protocolo de tratamiento bajo estudio, y c) identificación de la población elegible. Por ejemplo, si se desea poner en práctica una intervención para evitar la progresión de una enfermedad, deberían reclutarse sujetos que demuestren hallarse en fases tempranas de la enfermedad bajo estudio, sin incluir sujetos que están en riesgo de sufrirla.

Por estas razones, se han implementado diversas estrategias para evaluar la pertinencia en el planteamiento y reporte de ensayos clínicos aleatorizados. La publicación de un ECCA debe transmitir al lector, de manera clara, por qué se llevó a cabo el estudio y cuáles fueron los criterios de su conducción y análisis.[15] En 1981, un grupo de investigadores clínicos propusieron guías clínicas para usuarios de la literatura médica, con el fin de evaluar críticamente la información de artículos sobre diferentes tópicos, que incluyeran estudios sobre el tra-

**Cuadro III** **Lista de comprobación de puntos a revisar cuando se informe de un ensayo clínico controlado aleatorizado**

| *Sección y tema* | *Punto* | *Descriptor* |
|---|---|---|
| TÍTULO Y RESUMEN | 1 | Explicar cómo se asignan los participantes a las intervenciones. |
| INTRODUCCIÓN<br>Antecedentes | 2 | Proporcionar los antecedentes científicos, explicación y razonamiento. |
| MÉTODOS<br>Participantes | 3 | Determinar los criterios de selección de los participantes, así como la población y las localidades donde se recolectaron los datos. |
| Intervenciones | 4 | Precisar detalles de las intervenciones para cada grupo, y también cómo y cuándo fueron realmente administradas. |
| Objetivos | 5 | Especificar los objetivos y las hipótesis. |
| Resultados | 6 | Definir claramente las medidas primarias y secundarias de los resultados y, cuando aplique, cualquier método utilizado para incrementar la calidad de las mediciones (por ejemplo: múltiples observaciones, entrenamiento por asesores). |
| Tamaño de la muestra | 7 | Explicar cómo fue determinado el tamaño de la muestra y, cuando aplique, proporcionar la explicación de cualquier análisis intermedio y las reglas de terminación del estudio. |
| Aleatorización<br>Generación de la secuencia | 8 | Definir el método utilizado para implementar la secuencia aleatoria de asignación, incluyendo detalles de cualquier restricción (por ejemplo: bloques o estratificación). |
| Distribución a ciegas | 9 | Describir el método utilizado para implementar la secuencia aleatoria de asignación (por ejemplo: contenedores numerados o guía telefónica central, que aclare si la secuencia fue cegada hasta que las intervenciones fueron asignadas). |
| Implementación | 10 | Definir quién generó la secuencia de asignación, quien reclutó a los participantes y quien asignó a los participantes a su grupo. |
| Cegamiento (blindaje o enmascaramiento) | 11 | Explicar si los participantes, los que administraron la intervención y los que evaluaron los resultados fueron ciegos a la asignación de grupos. De haber sido así, definir cómo se evaluó el éxito del proceso de cegamiento. |
| Métodos estadísticos | 12 | Describir los métodos estadísticos utilizados para comparar los grupos en sus resultados primarios; métodos de análisis adicionales, tales como análisis de subgrupos o análisis ajustados. |

**Cuadro III** ***Continuación***

| *Sección y tema* | *Punto* | *Descriptor* |
|---|---|---|
| RESULTADOS<br>Flujo de participantes | 13 | Informar sobre el flujo de participantes a través de cada etapa (el uso del diagrama es fuertemente recomendado). Específicamente para cada grupo, notificar el número de participantes asignados en forma aleatoria, que recibieron el tratamiento establecido —pretendido—, completaron el protocolo de estudio y fueron analizados para los resultados primarios. Describir las desviaciones del protocolo de estudio diseñado, siempre junto con las razones. |
| Reclutamiento | 14 | Determinar las fechas definidas para los periodos de reclutamiento y seguimiento. |
| Datos basales | 15 | Proporcionar los datos demográficos basales y características clínicas de cada grupo. |
| Números analizados | 16 | Notificar el número de participantes (denominador) en cada grupo incluido en cada análisis, y si el análisis fue por "intención de tratar", explicar cuál es mejor. Presentar los resultados en números absolutos cuando sea posible (por ejemplo: 10/20; no 50%). |
| Resultados de la estimación | 17 | Para cada resultado primario y secundario, un resumen de resultados de cada grupo y el tamaño del efecto estimado de la muestra (por ejemplo: intervalo de confianza de 95%). |
| Análisis secundarios | 18 | Agregar multiplicidad y proporcionar información sobre cualquier otro análisis realizado, incluyendo análisis de subgrupos y análisis ajustados, e indicar cuáles fueron preespecificados y cuáles exploratorios. |
| Eventos adversos | 19 | Notificar todos los eventos adversos o efectos secundarios importantes, que se hayan presentado en cada grupo de intervención. |
| COMENTARIOS<br>Interpretación | 20 | Interpretar los resultados, tomando en cuenta las hipótesis de estudio, fuentes de sesgo potencial o imprecisión, así como los peligros de la multiplicidad de análisis y eventos de interés. |
| Generalización | 21 | Hacer una generalización (validez externa) de los hallazgos del estudio. |
| Evidencia global | 22 | Hacer una interpretación general de los resultados en el contexto de la evidencia actual. |

tamiento.[16] Por su parte, a mediados de 1990, el grupo para establecer las Normas para Información sobre Ensayos (SORT, por sus siglas en inglés) y el grupo de trabajo Asilomar, encargado de las recomendaciones para los informes de los ensayos clínicos en la literatura biomédica, desarrollaron en forma conjunta las Normas Consolidadas para la publicación de ensayos clínicos, las cuales se editaron en 1996.[17] Asimismo, la Declaración CONSORT comprende una lista de comprobación de 22 puntos y un diagrama de flujo para comunicar un ECCA. Por conveniencia, la lista de comprobación y el diagrama juntos se denomina sencillamente CONSORT, y ha sido diseñada principalmente para escribir, revisar y evaluar informes de ECCA simples de sólo dos grupos paralelos. El registro de un ECCA impli-

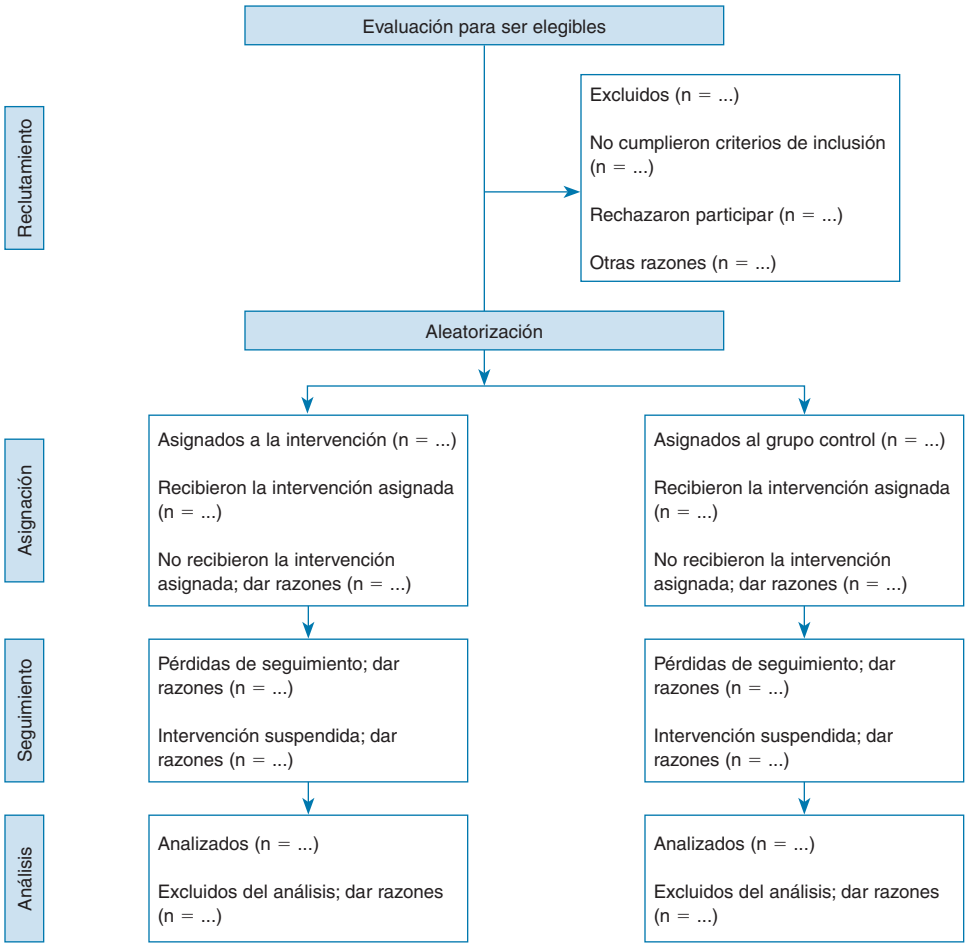



**Figura 5** **Diagrama de flujo del progreso por medio de las fases de un ECA**

ca revisar la lista de comprobación y el diagrama de flujo de CONSORT y también incluye puntos no esenciales pero altamente recomendables, como la aprobación del ensayo por un comité ético institucional, fuentes de financiamiento para el ensayo, y un número de registro para el ensayo [por ejemplo, el Número Internacional de Ensayo Aleatorio Controlado Estándar (International Standard Randomised Controlled Trial Number: ISRCTN)], utilizado para registrar el ECCA a su inicio. La lista y el diagrama de flujo pueden ser observados en el cuadro III y la figura 5.

### Blindaje

El cegamiento o blindaje es una maniobra que limita la ocurrencia de sesgos, conscientes o inconscientes, previene la identificación del tra-

**Cuadro IV Tipos y ventajas del blindaje o enmascaramiento**

| *Sujetos blindados* | *Beneficios potenciales* |
|---|---|
| Participantes | Ausencia de sesgo en las respuestas fisiológicas o físicas motivadas por el tipo de intervención. |
| | Mayor adherencia al régimen de tratamiento. |
| | Menor búsqueda de intervenciones adicionales adyuvantes. |
| | Menores pérdidas en el seguimiento. |
| Diseño | Ausencia de transferencia de inclinaciones o actitudes a los participantes. |
| Investigadores | Administración no diferencial de cointervenciones. |
| | Imposibilidad de ajustar dosis. |
| | Imposibilidad de asignación diferencial de los participantes. |
| | Imposibilidad de alentar o desalentar diferencialmente la adherencia al estudio. |
| Análisis | Ausencia de sesgos en la evaluación de resultados. |
| | *TIPOS DE BLINDAJE* |
| *Tipo de blindaje* | *Significado* |
| No ciego | Ensayo en el que tanto los investigadores como los participantes conocen el tratamiento asignado |
| Ciego | Indica que los participantes, investigadores del ensayo o patrocinadores desconocen la intervención asignada. |
| Simple ciego | Una de las tres categorías de individuos, normalmente participantes, desconoce el tratamiento asignado a lo largo del desarrollo del estudio. |
| Doble ciego | Participantes, investigadores y patrocinadores desconocen el tratamiento asignado. |
| Triple ciego | Ensayo doble ciego que mantiene blindado el análisis de los datos, hasta una etapa determinada del estudio. |

tamiento asignado, y el curso de tratamiento u observaciones previas.[18] Al desconocer el tipo de tratamiento, se garantiza que variables tales como reclutamiento de sujetos, cuidados subsecuentes, actitudes de los sujetos hacia su tratamiento, estimación de los resultados finales y exclusión de datos del análisis, no influyan en la adherencia al protocolo. Los procesos usualmente blindados son la intervención asignada y la evaluación del estatus de los sujetos de estudio. El blindaje previene determinados sesgos en las diversas etapas del ensayo clínico y protege la secuencia después de la asignación al grupo de tratamiento. Existen básicamente tres niveles de blindaje, entre los que se encuentran el simple, doble y triple, cuyas características se describen en el cuadro IV.

### Placebo

Un placebo es un agente diseñado para simular una terapia médica y responde a un procedimiento que no tiene efecto fisiológico ni bioquímico sobre la enfermedad o condición de estudio, aunque puede mostrar un perfil de

**Cuadro V Acuerdos internacionales sobre el uso de placebo**

| *Declaración de Helsinki (2000)*[a] |
|---|
| Los beneficios, riesgos, cargas y efectividad de la nueva terapia [o tratamiento], deberían ser evaluadas contra los mejores métodos profilácticos, diagnósticos y terapéuticos actuales. Esto no excluye el uso de placebo ni tratamiento en estudios donde no existe este estándar o tratamiento probado como el mejor. |
| Nota de aclaración 2002[b, c] Los ensayos clínicos con brazo placebo deberían ser aplicados solamente en ausencia de una terapia existente y probada, en las siguientes situaciones: a) por razones metodológicas y científicas para determinar la seguridad y eficacia de un método profiláctico, diagnóstico o terapéutico, b) si el método profiláctico, diagnóstico o tratamiento es investigado para una entidad clínica menor y los sujetos de la investigación no serán objeto de algún riesgo adicional o daño irreversible. |
| *CIOMS (2002)* [d, e] |
| El uso de placebo está indicado:<br>1. Cuando no existe una intervención de efectividad comprobada.<br>2. Cuando la omisión de una intervención de efectividad comprobada expondría a los sujetos, a lo sumo, a una molestia temporal o a un retraso en el alivio de sus síntomas, y<br>3. Cuando el uso de una intervención de efectividad comprobada como control no produciría resultados científicamente confiables, y el uso de placebo no añadiría ningún riesgo de daño serio o irreversible para los sujetos. |

a. World Medical Association (1996, 2000). Declaration of Helsinki. Edinburgh, Scotland. [párrafo 29].
b. World Medical Association. Declaration of Helsinki as amended by the 52nd WMA General Assembly, Edinburgh, Scotland, October 2000; Note of Clarification on Paragraph 29 added by the 53rd General Assembly of the WMA, Washington, DC, 2002.
c. Aquí se manifiestan dos críticas importantes; una aclaración éticamente aceptable incluiría que ambas condiciones se complementen en el momento de considerar el uso de placebo, y otro punto importante es que ambas condiciones prevendrían a los sujetos participantes de un daño serio e irreversible. Se entiende que minimizar daños constituye un requerimiento ético en investigación.
d. Council for International Organizations of Medical Sciences [CIOMS], 1993; 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland. Guideline No. 5.
e. Por regla general, los sujetos de una investigación en el grupo control de un ensayo de diagnóstico, terapia o prevención, deberían recibir una intervención de efectividad comprobada. En algunas circunstancias, puede ser éticamente aceptable usar un control alternativo, tal como placebo o "ausencia de tratamiento" [CIOMS, Pauta 11, 2002].

efectos secundarios similares al del fármaco o intervención estudiada. Este procedimiento es útil para mantener el desconocimiento de los pacientes y clínicos sobre la intervención asignada, además de crear un "control negativo" que difiere del grupo en estudio sólo por la intervención. De esta manera, es posible estimar el efecto real de la intervención, deduciéndolo del efecto logrado por el placebo en la misma situación clínica.

El primer ensayo clínico controlado con placebo fue probablemente el conducido en 1931, cuando se probó el *sanocrysin* en comparación con agua destilada en pacientes con tuberculosis.[19] Desde entonces, los ensayos clínicos aleatorizados con placebo han sido controvertidos, especialmente cuando los participantes son asignados de manera aleatoria en una de sus ramas (placebo, por ejemplo) y por lo tanto se les priva del tratamiento efectivo.[20,21] En los ensayos clínicos, el placebo se traduce en tratamientos de control con similar apariencia a los tratamientos en estudio, pero sin su actividad específica.[22] El uso de placebo ha reportado resultados objetivos y subjetivos en un rango de 30 a 40% de los pacientes, en una amplia gama de entidades clínicas, entre las que destacan dolor, asma, hipertensión arterial sistémica e, incluso, infarto al miocardio, entre otras.[23] Desde hace 50 años, se ha documentado que el uso de placebo tiene un alto grado de efectividad, interpretado bajo mecanismos desconocidos como un efecto terapéutico real en más de 30% de los casos.[24] Por esta razón, el efecto placebo no podría distinguirse de la historia natural de la enfermedad, regresión a la media o el efecto de otros factores.

El debate actual acerca del uso apropiado del placebo en las investigaciones clínicas surge después de la publicación de numerosos ejemplos de ensayos clínicos en los que se empleaba placebo, a pesar de la existencia de un tratamiento efectivo.[25] Dichos estudios violaban los principios básicos de la Declaración de Helsinki. Además, la aparición del VIH-SIDA y las innovadoras metodologías para evaluar drogas de reciente aparición (ensayos clínicos multinacionales, financiadores externos al país huésped, uso de comparadores —placebo, por ejemplo—), sobre todo en las naciones con menores ingresos, provocaron debates éticos por parte de los miembros de las instancias reguladoras internacionales (cuadro V).



## CONCEPTO Y MÉTODOS DE ALEATORIZACIÓN

La asignación aleatoria define y diferencia el ECCA de los estudios observacionales, porque es la única intervención metodológica que teóricamente da lugar a una distribución equilibrada de las características de los sujetos entre los diferentes grupos de intervención o tratamiento. Por lo tanto, garantiza la comparabilidad de las poblaciones y maximiza el principio de intercambiabilidad, al mininizar los sesgos de selección. El propósito primario de la aleatorización es, entonces, garantizar que la posible inferencia causal observada al final del estudio no se deba a otros factores. Se ha sugerido una gran variedad de procedimientos de aleatorización en la literatura. La aleatorización se refiere a la asignación mediante el azar de las unidades de investigación a uno de dos o más tratamientos, con la finalidad de compararlos sobre las variables de desenlace de interés. Se acepta que la aleatorización tiene como propósito prevenir la existencia de diferencias entre los grupos que no se deriven de los tratamientos que están en comparación. De esta manera, cuando se produce un equilibrio de las posibles variables que podrían modificar y determinar el efecto del tratamiento

sobre la variable de desenlace, las diferencias deben considerarse estrictamente en relación con la maniobra bajo estudio.

El concepto de aleatorización, originalmente fue utilizado por Fisher en su texto clásico "El diseño de experimentos", que tiene como argumento principal que la aleatorización prevendría las diferencias sistemáticas de cualquier tipo, independientemente de que el investigador pudiera identificarlas. Este concepto vuelve el proceso de aleatorización preferible sobre la asignación no probabilística (sistemática, secuencial, por facilidad o por conveniencia) la cual, en ningún momento, tiende a asegurar el equilibrio entre los grupos. Sin embargo, siempre hay que tener cuidado, ya que puede ser que la aleatorización no dé como resultado una distribución equilibrada de las variables confusoras entre los grupos. En consecuencia, es indispensable efectuar una comparación de las variables en el momento en que ingresa el paciente al estudio, mismas que podrían afectar los efectos de la maniobra, para que, en caso de que existan diferencias significativas, se ajusten los resultados obtenidos por dichas variables. A continuación, se describen ampliamente los métodos utilizados para asignar la maniobra de intervención.[26-37]

## Asignación aleatoria de los tratamientos

Un objetivo importante de la investigación clínica es el desarrollo de terapias que mejoren la probabilidad de desenlaces exitosos en los sujetos enfermos o que prevengan el inicio de la enfermedad en los individuos sanos. La evidencia convincente de la efectividad de una maniobra requiere no sólo de observar una diferencia entre los grupos respecto al desenlace de interés, sino de demostrar que la maniobra es la que más probablemente ha causado dichas diferencias. Por ejemplo, los pacientes que acaban de ser sometidos a una nueva intervención quirúrgica pueden presentar un mayor tiempo de supervivencia que los que recibieron la intervención convencional. Habría que analizar hasta qué grado el resultado de la supervivencia depende de la cirugía y no de la habilidad del cirujano para seleccionar a los pacientes de bajo riesgo quirúrgico. Para asegurar una evaluación no sesgada de los tratamientos, los grupos de estudio deben ser equivalentes en todo, excepto en las maniobras que reciben (*ceteris paribus*).

## Necesidad de aleatorización en los ensayos clínicos

En muchos experimentos realizados en el laboratorio, los científicos poseen las herramientas para lograr equivalencia entre las unidades bajo investigación; con capacidad para mantener un control perfecto sobre las muestras que se están comparando, experimentos pequeños pueden ser suficientes para medir en forma precisa los efectos de las maniobras. En biología sin embargo, especialmente cuando los sujetos bajo estudio son los seres humanos en su totalidad, la variabilidad inherente ocasiona que el control de los potenciales confusores sea prácticamente imposible. Si bien existen algunas estrategias para parear grupos asignados a las diferentes maniobras de acuerdo con los posibles confusores, esto generalmente requiere tamaños de la muestra elevados en cada estrato, lo que imposibilita hacer el balance, debido a la existencia de variables desconocidas y el estado actual del conocimiento.

Además de seleccionar grupos comparables, hay que tener cuidado que el médico no tenga preferencias conscientes o inconscientes por pacientes específicos, de acuerdo con el tratamiento a probar, ya que puede provocar sesgos en los resultados (esta es la razón por la cual se debe cegar al evaluador durante la medición). Por lo tanto, es importante que la asignación de los pacientes a los siguientes tratamientos sea aleatoria.

## Aleatorización como base para la inferencia estadística

De acuerdo con la teoría frecuentista, la aleatorización permite realizar pruebas directas de causa y efecto, así como construir pruebas válidas de significancia estadística. Por ejemplo, en un ensayo que evalúa el efecto de un nuevo medicamento sobre el nivel de LDL-Colesterol, el modelo se inicia justamente con la medición de los niveles basales de LDL-C de cada paciente. Por medio de un proceso de aleatorización, se les prescribe el nuevo medicamento de interés o la maniobra comparativa, y al final del periodo de estudio, se vuelven a medir los niveles. Si lo que se esperaba era que el medicamento redujera LDL-C en 10 mg/dL, entonces el grupo al que se le administró el nuevo medicamento deberá de tener en promedio niveles de LDL-C menores que el grupo control. Para probar este efecto se comparan los promedios observados en ambos grupos. Gracias al proceso de aleatorización, que en términos generales balancea los posibles confusores entre los grupos, se espera que si las diferencias son suficientemente grandes y significativas, puede concluirse que fue el tratamiento el que las produjo.



## Asignación aleatoria simple

Como se comentó líneas arriba, la aleatorización requiere un mecanismo gobernado por el azar para asignar las maniobras (los tratamientos) a los sujetos bajo investigación. Los ensayos clínicos reales deben utilizar métodos verificables, de tal manera que, después del estudio, el investigador pueda demostrar que la asignación se mantuvo libre de sesgo. La manera más sencilla de asignar la maniobra de intervención es la aleatorización simple. En ella se utiliza como herramienta base el cuadro de números aleatorios. Se selecciona al azar un punto de inicio y posteriormente se selecciona la dirección de movimiento que se mantendrá constante a lo largo de todo el cuadro. Se decide a priori qué grupos de números (0 al 9) se destinarán a cada maniobra y, por ejemplo, puede convenirse que los números pares (0, 2, 4, 6 y 8) se destinen a la maniobra A y los nones (1, 3, 5, 7 y 9) a la B. En el cuadro VI se incluye una sección de una serie de números aleatorios contenida en cualquier libro de bioestadística básica, en la que se seleccionó por azar el punto de inicio en el segundo renglón de la segunda fila y en el primer número 5 que aparece. En este caso se selecciona en dirección de izquierda a derecha y arriba abajo, de tal forma que, partiendo del 5, los números que continúan son el 6, 8, 9, 7 y 2. De esta manera, dado que el primer número a emplear es el 5, el primer sujeto que ingrese al estudio recibirá la maniobra B, el segundo la maniobra A, el tercero la maniobra A, el cuarto la maniobra B y así sucesivamente hasta asignar el total de sujetos necesarios a incluir.

Debido a lo tedioso que puede resultar la asignación manual cuando se trata de una gran cantidad de sujetos, y a que, de acuerdo con algunos investigadores, esta forma puede motivar algunos sesgos, la lista de aleatorización debe generarla una computadora, siempre que sea posible. La persona que trabaja en el equipo de cómputo para generar la lista de aleatorización debe ser ajena a las personas que reclutan y valoran a los participantes de la investigación. Durante el curso del estudio, el generador de las listas no debe divulgar los detalles del método particular utilizado para generar las listas.

Muchos estudios utilizan sobres opacos, en cuyo interior se encuentra el siguiente tratamiento por asignar. Este método, aun cuando es considerado por muchos investigadores como estándar, puede ser sujeto a violación, especialmente en estudios no cegados. Un investigador que desea que determinado tipo de pacientes ingresen a una rama específica del estudio puede observar a contraluz el contenido del sobre o, incluso, abrirlo y volverlo a cerrar.

**Cuadro VI** **Ejemplo de aleatorización simple, por medio de la tabla de números aleatorios**

| *Serie de números* | | *Paciente* | *Maniobra* |
|---|---|---|---|
| 8467893 | 5489631 | 1 | B |
| 0236792 | 4568972 | 2 | A |
| 2467810 | 1348392 | 3 | A |
| 3112348 | 3476812 | 4 | B |
| 5912902 | 0981345 | 5 | B |
| 7645690 | 3289732 | 6 | A |
| | | | |
| 5674389 | 2310398 | 7 | A |
| 2938001 | 3289923 | 8 | A |
| 1345698 | 4728625 | 9 | A |
| 3298567 | 1223938 | 10 | B |
| 3490594 | 1309093 | 11 | A |
| 5489207 | 4532904 | 12 | B |

Lo anterior ha generado que muchos estudios en la actualidad adopten códigos de aleatorización generados vía telefónica, por fax o encriptados, por medio de sistemas de cómputo.

Una de las desventajas que tiene este tipo de aleatorización es que, cuando las muestras son pequeñas, por lo general se producen desbalances en el número de sujetos asignados a cada tratamiento, ya que puede asignarse un mayor número de sujetos a determinada maniobra. Otra limitante importante es que, en ocasiones, se producen secuencias repetidas de una misma maniobra (sujetos 4, 5 y 6 a la A; luego sujetos 7 y 8 a la B; después sujetos 9, 10, 11,12 y 13 a la A, y así sucesivamente), ya que los sujetos que ingresan en un momento determinado al estudio pueden ser distintos en sus características basales o en la forma de responder a las maniobras.

### Aleatorización en bloques balanceados

Tratando de limitar la posibilidad de desbalances en la asignación de tratamientos, de generar secuencias repetidas largas de una misma maniobra y de balancear en la medida de lo posible algunos de los sesgos inherentes al proceso de aleatorización simple, se creó el método de aleatorización en bloques balanceados. Se ensambla una serie de bloques, formados por un número determinado de celdas, en las cuales se incluyen los distintos tipos de tratamiento. El número de bloques estará determinado por el número de participantes a incluir en el estudio. Cada bloque contendrá en cada celda una de las alternativas de tratamiento y dentro de cada bloque deberá existir un numero balanceado de los posibles tratamientos (cuando, por fines éticos y de seguridad, se considera conveniente asignar el doble o triple de pacientes a una determinada maniobra, se dice que se trata de una aleatorización en bloques desbalanceados, ya que existirá el doble o triple de celdas en cada bloque de uno de los tratamientos).

En el cuadro VII se presenta un caso de aleatorización en bloques balanceados, en el que se asignaron 24 sujetos a dos alternativas de tratamiento, con bloques balanceados con longitud fija de cuatro celdas por bloque. Dado que se trata de 24 pacientes y se incluirán cuatro celdas en cada bloque, se necesitarán seis bloques (número de bloques = número de pacientes entre el número de celdas por bloque). Dado que se incluirán cuatro celdas por bloque, y sólo existen dos alternativas de tratamiento, en cada bloque deberán incluirse las diferentes combinaciones de A y B. Quien asigna el número de uso a cada bloque es el cuadro de números aleatorios. Así que, por azar, el primer número del tercer renglón de la primera columna es el número 2, por lo tanto, el primer bloque es el número 2; el segundo es el 4; el tercero es el 6 en uso, y luego, debido a que los números 7 y 8 no se utilizan, el cuarto es el 1; luego el cero no se usa y, si se decide continuar en la dirección marcada con la flecha, los números siguientes son el 1 (repetido) y el 3 para el quinto; finalmente, el sexto bloque es

**Cuadro VII Ejemplo de aleatorización simple, por medio de la tabla de números aleatorios**

| *Serie de números* | | *2* | *4* | *6* | *1* | *3* | *5* |
|---|---|---|---|---|---|---|---|
| 8467893 | 5489631 | A | B | A | B | A | B |
| 0236792 | 4568972 | A | B | B | A | B | A |
| 2467810 | 1348392 | B | A | A | B | B | A |
| 3112348 | 3476812 | B | A | B | A | A | B |
| 5912902 | 0981345 | | | | | | |
| 7645690 | 3289732 | | | | | | |
| | | Pacientes | | | | | |
| 5674389 | 2310398 | 1. B | 5. A | 9. | 13. | 17. | 21. |
| 2938001 | 3289923 | | | | | | |
| 1345698 | 4728625 | 2. A | 6. A | 10. | 14. | 18. | 22. |
| 3298567 | 1223938 | 3. B | 7. B | 11. | 15. | 19. | 23. |
| 3490594 | 1309093 | | | | | | |
| 5489207 | 4532904 | 4. A | 8. B | 12. | 16. | 20. | 24. |

el 5. Una vez asignado el número a cada bloque, se utilizan las combinaciones de tratamientos contenidas dentro de ellos.

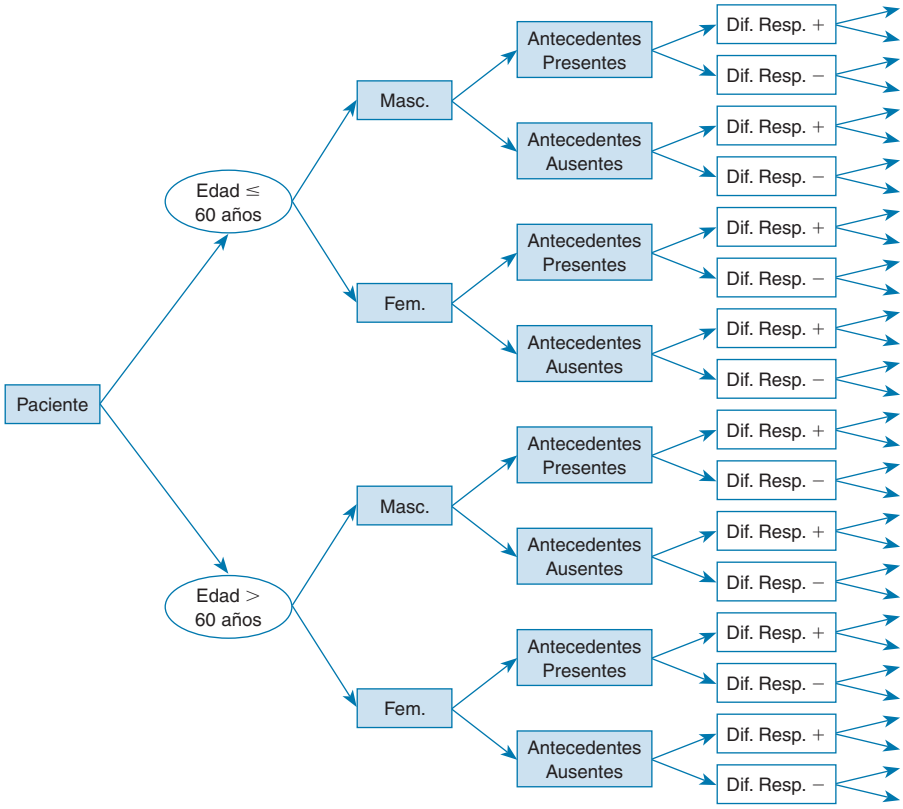
### Aleatorización estratificada

Si bien la aleatorización en bloques balanceados permite la asignación equilibrada de los sujetos a las maniobras, independientemente del momento en que se decida detener el ensayo y del número de pacientes incluidos hasta ese punto, tiene la desventaja de que no permite efectuar un balanceo por las posibles variables confusoras. Esto ha llevado a que diversos investigadores señalen que, por medio de la revisión sistemática o narrativa de la bibliografía y de la consulta con expertos o de la experiencia misma, es preferible identificar a priori los factores que podrían modificar el impacto de la maniobra sobre la variable de desenlace para, de esta manera y dependiendo de la factibilidad en cuanto al tamaño de la muestra, decidir cuántos estratos establecer a la asignación de la maniobra (y en cuántos se efectuará ajuste de los resultados en el momento del análisis, si es que quedan desbalanceados después de la asignación aleatoria). El cuadro VIII tiene un ejercicio en el que se determinaron como potenciales modificadores de los efectos de la maniobra sobre la variable de desenlace una edad mayor de 60 años, el género masculino, la historia familiar y la presencia de dificultad respiratoria en el momento de ingreso al estudio. Se establecen los diferentes subestratos y al final de la estratificación se efectúa la asignación de la maniobra por medio de asignación simple o en bloques balanceados (dependiendo del número de pacientes disponibles y de la posibilidad de desbalances en la asignación); puede observarse cómo inmediatamente este método degenera conforme aumenta el número de estratos (por ejemplo, cuando se tienen cinco posibles confusores con tres categorías para cada uno, se tendrá la necesidad de ensamblar $3^5 = 243$ distintos estratos).

### Aleatorización en conglomerados (grupos o "clusters")

Se trata de un proceso de aleatorización simple o en bloque de grupos de personas, salo-

**Cuadro VIII** **Ejemplo de aleatorización estratificada**




nes, delegaciones, comunidades, municipios, ciudades, estados o países. La unidad de asignación es el grupo y no el individuo. Este tipo de aleatorización es ampliamente utilizada en investigaciones epidemiológicas, aunque también puede usarse para evaluar el impacto de programas sociales, educativos y medidas preventivas comunitarias, entre otros casos.

No obstante que, para fines de aleatorización, la unidad de interés es el grupo y no los individuos que lo conforman, es importante medir el grado de similitud de respuesta dentro del conglomerado. Para ello se utiliza el cálculo del coeficiente de correlación intraclase o intraconglomerado, denotado por la letra griega ρ. Este parámetro puede interpretarse como el coeficiente de correlación estándar de Pearson entre cualesquiera de dos respuestas en el mismo conglomerado, y puede ser calificado como positivo alto, positivo bajo, cero, negativo bajo, y negativo alto. En términos generales, cuando el valor de ρ es positivo, se asume que la variación entre las observaciones en diferentes conglomerados excede la variación dentro de los conglomerados. Las razones pueden deberse a la manera de seleccionar los sujetos; a la influencia de covariación del conglomerado; la posibilidad de compartir algunos factores mas rápidamente dentro de él, que entre las comunidades, o el efecto de las

interacciones personales entre los miembros que reciben la misma maniobra.

### Asignaciones dinámicas o adaptativas de los tratamientos

Los diseños más simples de los ensayos clínicos controlados incluyen un predeterminado número de individuos bajo estudio que, por medio de los métodos tradicionales ya descritos, se asignan a alguno de los tratamientos de interés con igual probabilidad para cada paciente de recibir una u otra modalidad terapéutica. En forma cada vez más frecuente, tanto la manera en que los sujetos de estudio son asignados a los tratamientos, como la forma en la que se decide terminar el estudio, se basan en la información que se va generando durante el progreso del estudio. Existen dos formas de asignación de los tratamientos: dinámica y adaptativa. La primera se aplica cuando la información sobre los covariados del paciente que predicen el desenlace clínico se utiliza para determinar la asignación del tratamiento; la segunda, cuando se utilizan datos de desenlace acumulados que afectan la selección del tratamiento.

La asignación dinámica permite efectuar un balance individual de los posibles confusores, sin tener que efectuar un balance dentro de todas las combinaciones de los factores (como en el caso de la aleatorización estratificada tradicional). Supóngase que existen $f$ factores y $l_f$ niveles en el factor $f$. En cualquier momento dado durante el ensayo, la asignación del tratamiento del paciente previo habrá creado algo de desbalance entre los factores. Si dejamos que sea $t_{ijk}$ el número total de pacientes en el $j$th nivel del factor $i$, que ha sido asignado al tratamiento $k$, $i$=1, …,$f$, $j$=1,…,$lf$, $k$=1,…,$r$ (donde $r$ es el número de tratamiento), el ensayo está balanceado para el factor $i$ nivel $j$, al grado de que $tij$1,…,$tijr$ son similares. Si el siguiente paciente a ser aleatorizado posee el factor $i$ al nivel $j$th, entonces uno puede considerar el efecto que cada posible asignación de tratamiento tendría en este balance. Dado que el balance debe ser caracterizado por una función matemática, se ha propuesto un método de minimización, donde el balance se caracteriza por un rango de tratamientos totales, y el tratamiento se selecciona por minimización de la suma a través de todos los factores. Existe también una versión más general de este método en el cual el tratamiento se selecciona por medio de una aleatorización por moneda sesgada, con las probabilidades de la moneda sesgada determinadas por la función de balanceo. La función de balanceo general involucra una suma ponderada de las funciones de balanceo de los factores individuales, donde los pesos podrían ser asignados sobre la base de la importancia relativa de cada factor confusor.

Por su parte, los diseños adaptativos, los cuales dependen de la acumulación de datos de desenlace, se han descrito desde principios de 1950. Armitage, por ejemplo, propuso un esquema que permitía la terminación global del estudio basado en un nivel de significancia global, lo que dio origen a los famosos análisis intermedios (análisis *interim*), en los que se producía una culminación prematura del estudio cuando se alcanzaban ciertos límites de desenlace y significancia estocástica. En 1963, estas ideas fueron revolucionadas por Colton, quien propuso reglas de detención de los ensayos basadas en pérdidas de la función adecuadas, en contraste con las reglas estocásticas utilizadas previamente (de significancia estadística). Para ello se propuso la construcción de un "horizonte del paciente", en el que los pacientes se aleatorizaban hasta que se cruzaban los límites preestablecidos, después de lo cual todos los pacientes restantes se asignaban al tratamiento con la mayor eficacia. Así, los límites óptimos se establecían al intercambiar las pérdidas generadas por aleatorizar la mitad de los pacientes al tratamiento inferior.

En 1969, Zelen popularizó este método con el nombre de sistema "Jugando al ganador", en el cual el primer paciente se asignaba, por ejemplo, a la maniobra A; si se obtenía un éxito, el siguiente paciente era asignando al grupo A, hasta que se presentara una falla, y entonces el siguiente paciente se asignaba al B, y así sucesivamente. Esta evaluación secuencial tiene también la ganancia ética de que no expone al grupo placebo a riesgo innecesario para continuar recibiendo un placebo o un tratamiento inefectivo.

## ANÁLISIS DE UN ENSAYO CLÍNICO ALEATORIZADO

"Un ensayo clínico apropiadamente planeado y ejecutado es una técnica experimental poderosa para estimar la efectividad de una intervención."[38] Este concepto ha sido aplicado en numerosos estudios realizados en todo el mundo, bajo la premisa de que todo ensayo clínico controlado comienza con una planeación cuidadosa, y pasa por un proceso detallado de ejecución y monitoreo, sin menospreciar cualquier procedimiento por simple que parezca para garantizar la comparabilidad de los datos obtenidos. La estimación de los resultados se realiza mediante diferentes técnicas estadísticas que comentaremos más adelante; sin embargo, es imprescindible mencionar que la piedra angular de un análisis estadístico axiomático está fundamentada en el meticuloso planteamiento del diseño.

### Principio analítico de "intención de tratar"

Los ensayos clínicos aleatorizados se analizan por un método estándar llamado "intención de tratar". Este proceso consiste en analizar a todos los sujetos aleatorizados de acuerdo con la asignación original del tratamiento y todos los eventos contados contra el tratamiento asignado.[39] Con base en este principio, todos los estudios aleatorizados deberían analizarse bajo este concepto, ya que el análisis apoyado por la aleatorización mantiene la comparabilidad por medio de los grupos de intervención. Si el análisis excluye participantes después del procedimiento de aleatorización (ya sea porque no reciben el tratamiento originalmente asignado o muere antes de que se le dé el tratamiento) puede introducirse un sesgo, ya que los grupos de intervención se verán afectados por la falla en la aleatorización, el tamaño de la muestra y el blindaje de los grupos.[40]

Dos de las razones más comunes para que pacientes aleatorizados sean excluidos son la no elegibilidad posterior al procedimiento aleatorio. Citemos el trabajo del Anturane Reinfarction Trial Research Group,[41,42] que evaluó el efecto de la sulfinpirazona en pacientes que habían sufrido un infarto. La aleatorización incluyó en dos grupos a 1 629 pacientes, quienes sobrevivieron a un infarto del miocardio, uno tratado con el medicamento de prueba y otro con placebo. Los pacientes clasificados como elegibles fueron 1 558, mientras que 71 no reunieron los criterios de elegibilidad del protocolo. El análisis reportado se enfocó solamente en los pacientes elegibles, y mostró un efecto benéfico de la sulfinpirazona sobre la mortalidad, comparado con los no elegibles; sin embargo, un análisis posterior realizado por otros investigadores[43] reportó que la inclusión de aquellos no elegibles modificaba completamente los resultados, pues mostró que la interpretación de los resultados estaba sesgada por la exclusión de los sujetos declarados como no elegibles posterior a la aleatorización. Los resultados de este estudio fueron cuestionados por una agencia de regulación federal estadunidense.

Otra de las razones para la exclusión de pacientes en el análisis primario aparece cuando los pacientes no cumplen apropiadamente con la intervención especificada en el protocolo (falta de adherencia al tratamiento). Citemos el trabajo del Coronary Drug Project,[44] cuyo objetivo fue evaluar una estrategia medicamentosa para reducir el colesterol en hombres que sobrevivieron a un infarto del miocardio. El estudio tuvo dos grupos, uno tratado con bezofibrato y otro con placebo. Los resultados globales no mostraron diferencias en la mortalidad. Sin embargo, un análisis adicional clasificó en dos grupos a los pacientes: uno de ellos como buenos cumplidores del tratamiento (definido como que tomó sus medicamentos más de 80%) y otro grupo como poco cumplidor (<80%). La comparación de los poco cumplidores mostró una reducción de la mortalidad de 24.6% a 15%; sin embargo se observó una reducción mayor en la mortalidad en el grupo placebo (28.2% *vs.* 15.1%). Los buenos cumplidores vivieron más que los poco cumplidores independientemente del tratamiento. Estos resultados no pudieron explicarse por modelos de regresión, y los autores concluyeron que la adherencia a una intervención es, en sí misma, un resultado.

Aun cuando las exclusiones se prevean con anticipación en el protocolo, deberá tomarse en cuenta un posible sesgo, ya que los sujetos excluidos pueden diferir de los sujetos analizados en aquellas características medidas y no medidas, las cuales no siempre pueden probarse estadísticamente, debido a que la no adherencia al tratamiento no es un fenómeno aleatorio. Ningún método de análisis puede completamente contabilizar por el gran número de sujetos de estudio que se desviaron del protocolo de estudio, resultando en altas tasas de no adherencia, abandono o datos faltantes.

Se han hecho algunas recomendaciones para ajustar el diseño y aumentar el tamaño de la muestra que pueden compensar la falta de adherencia. En el diseño. debe elegirse un método de intervención con mejor tolerancia, y plantearse estrategias de supervisión durante la toma del tratamiento. Otra estrategia es aumentar el tamaño de la muestra para compensar el efecto de la falta de cumplimiento de los tratamientos. Se requiere un incremento de 23% en el tamaño de la muestra para compensar 10% de la falta de adherencia; sin embargo, para compensar 20% de la falta de adherencia, se requerirá un incremento de 56% del tamaño de muestra.[45]

El principio de análisis por "intención de tratar" es, por lo tanto, el método más conservador y provee una estimación de la efectividad del tratamiento (efecto del tratamiento dado a cada participante) más que una prueba verdadera de la eficacia del tratamiento (esto es, efecto del tratamiento en aquellos que siguieron el protocolo de estudio).

**Pérdidas en el seguimiento**

Aunque un estudio esté apropiadamente planeado para tener un tamaño de muestra capaz de compensar las pérdidas en el seguimiento, cuando éstas ocurren, el estudio puede sufrir una disminución del poder estadístico y aumentar la posibilidad de introducir un sesgo, en especial cuando la pérdida no es resultado del azar. Los pacientes más enfermos pueden no estar disponibles para regresar a la clínica a realizarse exámenes y mediciones clave para el resultado del estudio (se puede suponer que la aleatorización permitió que los pacientes graves quedaran igualmente distribuidos entre los grupos). En otros casos, los pacientes sometidos a toxicidad medicamentosa pueden estar indispuestos para ser evaluados en las mediciones clave. Otro punto importante que puede sesgar los resultados sucede cuando un grupo está más relacionado con los efectos adversos del medicamento, entonces la pérdida entre los grupos hará que éstos no sean comparables. Cuando éstos son los casos, los resulta-

dos pueden estar sesgados y ningún método estadístico podrá hacer ajustes, debido a la falta de datos. Entonces, se requerirán esfuerzos adicionales para localizar y estimar la variable de resultado primaria en aquellos participantes que no recibieron la intervención.

**Análisis por protocolo**

El análisis por protocolo es un método que incluye solamente un subgrupo de sujetos que cumplieron suficientemente con el protocolo, por lo que contrasta con el análisis conservador por "intención de tratar". El cumplimiento del protocolo incluye el cumplimiento de exposición mínima preespecificada al régimen de tratamiento, disponibilidad de mediciones de la variable primaria, elegibilidad correcta y ausencia de cualquier otra violación mayor al protocolo.[46] Excluye también eventos que ocurrieron después de que el sujeto dejó de adherirse al protocolo. Este método también se identifica como de "casos válidos", muestra de "eficacia" o de "sujetos evaluables".

Puede maximizar la oportunidad para un nuevo tratamiento que muestre eficacia adicional y refleje más cercanamente el modelo científico fundamental. Sin embargo, la correspondiente prueba de hipótesis y estimación del efecto del tratamiento puede o no ser conservadora, dependiendo del ensayo; el sesgo, que puede ser severo, emana del hecho de que la adherencia al protocolo puede estar relacionada con el tratamiento y, por ende, con el resultado.

Las razones precisas para la exclusión de los sujetos del grupo por protocolo deben ser cuidadosamente definidas y documentadas antes de romper el blindaje en una forma apropiada dentro de las circunstancias del ensayo específico. Los problemas que guían la exclusión de los sujetos para crear el subgrupo por protocolo, y otras violaciones al protocolo de estudio, deben identificarse y analizarse cuidadosamente. Las violaciones relevantes al protocolo pueden incluir errores en la asignación del tratamiento, uso de medicamentos excluidos, pobre cumplimiento, pérdidas en el seguimiento y datos faltantes. Es una buena práctica estimar el patrón de tales problemas entre los grupos de tratamiento respecto a la frecuencia y tiempo de ocurrencia.

**Resultados de variables subrogadas**

Debido a que las variables clínicas que son evaluadas por largo tiempo y aumentan los costos se usan como resultados en muchos ensayos clínicos, algunos investigadores han buscado alternativas para obtener ensayos clínicos más cortos y más pequeños por medio del uso de una variable respuesta sustituta del resultado clínico. Un requerimiento para obtener esta variable sustituta es que debe ser predictiva del resultado clínico. Otro requerimiento es que la variable sustituta debe capturar el efecto total de la intervención sobre el resultado clínicamente relevante. Aunque teóricamente posible, biológicamente es un fenómeno complejo donde la intervención puede modificar la variable subrogada y no tener efecto o tenerlo parcialmente sobre la variable clínica de respuesta. Por el contrario, la intervención puede modificar la variable clínica sin afectar a la subrogada.

Aunque se han propuesto criterios estadísticos para validar las variables subrogadas, la experiencia con su uso es limitado. En la práctica, la fuerza de la evidencia para la sustitución depende de tres aspectos: a) plausibilidad biológica de la relación entre la variable clínica y la variable subrogada; b) demostración en estudios epidemiológicos del valor pronóstico para la variable clínica, y c) evidencia de ensayos clínicos de que el efecto del tratamiento sobre la variable sustituta corresponde al efecto sobre la variable clínica. La interpretación de un ensayo que utiliza como base una variable subrogada, debe hacerse en el contexto del número de individuos en riesgo, y considerar las conclusiones con mucha precaución.

## Análisis de resultados múltiples y análisis por subgrupo

### *Análisis de resultados múltiples*

Aunque en los ensayos clínicos, comúnmente, se usa un evento único como variable respuesta primaria, en muchos otros la variable respuesta es una combinación de diversos eventos o se requieren múltiples mediciones de condiciones inherentes al tratamiento de prueba (múltiples comparaciones de tratamientos, mediciones repetidas, análisis por intervalos, etc.). Aquí se incluyen desde características clínicas tales como morbilidad, síntomas, calidad de vida, efectos colaterales del tratamiento, hasta utilización y costos de los servicios de salud. Es muy importante mencionar en el protocolo la especificación de la prioridad que tendrán las respuestas múltiples.

Tradicionalmente, se recomienda que sea diseñada en el protocolo una variable resultado primaria y las otras variables definirlas como secundarias o terciarias. El ensayo es bajo estas condiciones, ponderado y vigilado sobre las bases de una variable de resultado primaria. (Una variable primaria global sería la mortalidad por todas las causas vasculares, las cuales se analizarían separadamente como variables secundarias, los eventos que las formaron: infarto al miocardio, infarto cerebral, trombosis mesentérica, etc.). El agrupamiento de los eventos debe ser acorde con los objetivos del diseño para no mezclar eventos de mortalidad, morbilidad o calidad de vida.

Si se utilizan múltiples resultados primarios, se requiere de un número mayor de métodos estadísticos para contabilizar la multiplicidad y mantener al mismo tiempo los niveles de error tipo I en 0.05[47] (esto quiere decir que se requieren ajustes del nivel $\alpha$, valor de $p$, valores críticos para multiplicidad y pruebas globales que producen un índice resumido de la combinación de las variables resultado). Un método ampliamente usado es el de Bonferroni, cuyo nivel de significancia para cada variable resultado es $\alpha/k$, donde $k$ es igual al número de variables resultado; pero esta función es limitada si las variables resultado están correlacionadas. Se han hecho otras propuestas para evitar este tipo de problemas, como ordenar los valores univariados de $p$ de menor a mayor, y cada valor de $p$ se compara con los niveles de $\alpha$ progresivamente mayores desde $\alpha/k$, $\alpha/(k-1)$, hasta que $\alpha$ no rechace la hipótesis nula.[48]

Se han desarrollado otros métodos de pruebas para ajustar el valor crítico de multiplicidad que toman en consideración la distribución conjunta de las pruebas estadísticas de las múltiples variables resultado.[49] Sin embargo, Mantel[50] sugiere un ajuste menos conservador: $1-(1-\alpha)$, y Tukey[51] reemplaza $1/k$ con $1/\sqrt{k}$ cuando las variables resultado están correlacionadas, pero se desconoce su grado de correlación. Pueden realizarse estimaciones globales de la medición por medio de los métodos de O'Brien,[52] quien incluye métodos de suma de rangos y de regresión.

Otros métodos para analizar resultados de múltiples variables son el uso de componentes principales y de análisis factorial.[53]

### *Análisis por subgrupo*

Frecuentemente, la variable primaria se relaciona con otras variables de interés, en las que puede ser importante evaluar el efecto del tratamiento en el interior de subgrupos de sujetos. Tales variables pueden ser la edad, el sexo, o por qué razón los sujetos fueron tratados en diferentes centros (ensayos multicéntricos). Otra razón podría ser la consistencia interna de los resultados entre las diferentes categorías. Cualquiera que fuere la razón, es común dividir a los grupos aleatorizados en categorías o subgrupos para realizar comparaciones de la intervención en su interior. Es preferible que el establecimiento de los subgrupos se haya definido por anticipado en el protocolo, e identi-

ficado las variables que podrían influir sobre el resultado primario, ya que esto permite que puedan considerarse en el análisis (el prestablecimiento de los subgrupos y la preidentificación de variables) como una herramienta para mejorar la precisión y compensar cualquier falla en el balance de los subgrupos de tratamiento.

La confiabilidad de este análisis es pobre, debido a problemas de multiciplicidad y a que los grupos generalmente son pequeños en comparación con la muestra completa; por este motivo, el número de subgrupos debe mantenerse en un mínimo posible y especificarse previamente en el protocolo como medida de incremento en la credibilidad de los resultados, pues el manejo de la hipótesis refleja un conocimiento biológico a priori más que una argumentación de resultados a posteriori. Por lo anteriormente expuesto, los resultados de un análisis de subgrupos deben interpretarse como un análisis exploratorio, generador de hipótesis y nunca como un resultado definitivo. Estas recomendaciones, aparentemente fáciles de extrapolarse a los diseños, no se han aceptado del todo y se sigue cayendo en el riesgo de presentar interpretaciones inapropiadas de los resultados, como lo muestran Brookes y colaboradores[54] en su estudio de simulación para cuantificar el riesgo de mala interpretación de resultados, en este tipo de análisis. Los autores encontraron que, con un poder nominal de 80%, sólo tuvieron la oportunidad de detectar 29% de interacción en el efecto global del estudio. En el interior de los grupos sólo se detectaba entre 7% y 64%, lo que lleva a interpretaciones erróneas de la efectividad subgrupal de un tratamiento y no en su efecto global (mayor riesgo de falsos negativos). En algunos casos, para detectar la interacción con un poder significativo, tuvieron que inflar el tamaño de la muestra cuatro veces más. En este análisis, las técnicas estadísticas más utilizadas comprenden regresión lineal, árboles de regresión y estimaciones Bayesianas, cuando es necesario hacer un ajuste de sesgos.



## INTERPRETACIÓN DE RESULTADOS POR TIPO DE COMPARACIÓN

### Estudios de superioridad

En un ensayo clínico, la eficacia es el criterio que demuestra la superioridad de un tratamiento activo sobre un placebo, ya sea porque produce resultados claramente mejores o porque muestra un efecto dosis-respuesta. Los ensayos de superioridad intentan demostrar que la intervención experimental es mejor o igual que la intervención de control. Por ejemplo, para enfermedades graves, cuando un tratamiento ha demostrado ser eficaz por estudios de superioridad, puede considerarse una falta de ética proponer un nuevo ensayo clínico controlado con placebo. La nueva terapia debe ser más fácil de administrar, mejor tolerada, menos tóxica o menos costosa; pero el beneficio intercambiado no debe reflejarse en renunciar al efecto del tratamiento.

### Estudios de no inferioridad

En contraste con los estudios de superioridad, estos ensayos pretenden mostrar que la intervención experimental no es peor que el tratamiento estándar con algún margen de indiferencia. Adicionalmente, los investigadores deben demostrar que el nuevo tratamiento es mejor que el placebo en caso de que el ensayo tenga un brazo placebo. La conducción de un análisis de este tipo debe ser de muy alta calidad para detectar diferencias significativas; asimismo debe tener un efectivo control de la intervención, sin olvidar que el margen de in-

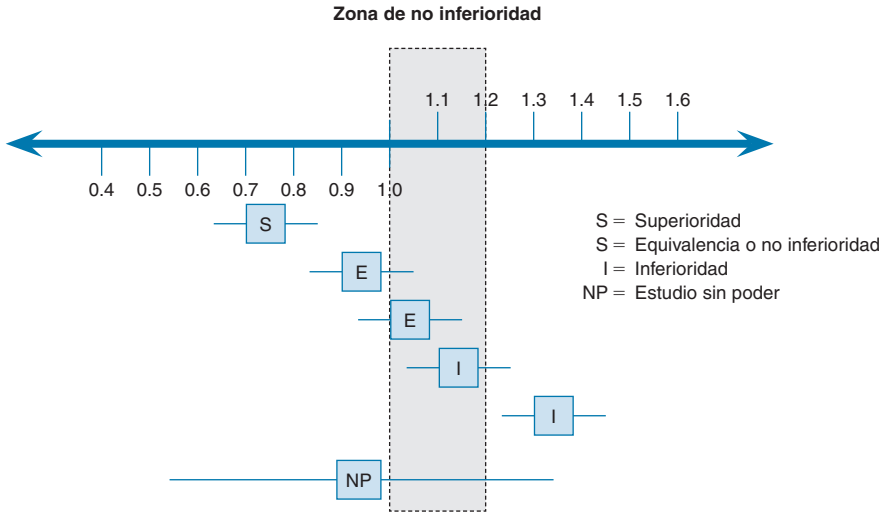



**Figura 6** **Zona de no inferioridad de los ensayos clínicos con este diseño (modificado de DeMets)**



diferencia no se basa en criterios estadísticos, sino en la importancia médica del tratamiento en el contexto de intercambio riesgo-beneficio. El análisis estadístico de estos estudios se basa generalmente en el uso de los intervalos de confianza (figura 6).

Citemos el caso del estudio OPTIMAAL,[55] que comparó losartan con captopril en pacientes con insuficiencia cardiaca. Sobre la base de que el primero podría tolerarse mejor que el segundo, su objetivo fue demostrar que losartan sería superior a captopril o, por lo menos, que no sería inferior. La mortalidad resultó en 1.126 (IC95% superior de 1.28), así que no se obtuvo como resultado ni la superioridad ni la no inferioridad. Los investigadores entonces usaron datos históricos de pacientes tratados con captopril (RR=0.806 comparado con placebo) y lo multiplicaron con los resultados de OPTIMAAL (1.126*0.806), y se aproximaron al efecto de losartan con placebo, si éste hubiera sido usado (el RR fue de 0.906); sin embargo, los resultados tuvieron que interpretarse con extrema precaución, a pesar de que el estudio OPTIMAAL estuvo muy bien conducido.

## MÉTODOS ESTADÍSTICOS USADOS EN LOS ECCA

Se han utilizado diversos diseños de análisis para evaluar las diferencias existentes entre los grupos sometidos a un tratamiento nuevo para compararlo con un placebo o un tratamiento convencional. La selección apropiada de análisis y la preespecificación en el protocolo del estudio es un reto que muchos investigadores no toman en cuenta. Perrone[56] llevó a cabo una investigación en 145 ensayos clínicos fase II en cáncer de mama, y encontró resultados inte-

resantes: 64.8% no tenía una propuesta estadística identificable, y los predictores más importantes para que identificaran un método estadístico definido (en modelos de regresión logística) el apoyo de una agencia financiadora, un único agente experimental y que el estudio fuera desarrollado como multicéntrico.

Los métodos de análisis estadístico han evolucionado en años recientes, desde el modelo tradicional de análisis de supervivencia con sus estimaciones, hasta llegar a modelos más complejos como los estudiados en los incisos previos, sin olvidarnos del análisis por riesgos competitivos (no censura el evento, como, por ejemplo, la muerte por otras causas, sino que toma en consideración su incidencia acumulada).[57] Sin embargo, en este documento daremos una perspectiva de las técnicas estadísticas más usadas en el desarrollo de los ensayos clínicos.

## Descripción de los grupos de comparación

Un primer análisis que se efectúa en los ensayos clínicos controlados consiste en comparar los grupos de tratamiento en sus condiciones basales —y que generalmente son características demográficas (edad, género), medidas antropométricas (masa corporal), condiciones inherentes a su estado clínico (severidad de enfermedad, control metabólico)— y otras variables pronósticas relacionadas con la variable resultado primaria. Generalmente se presentan como medidas de resumen (medias, modas, medianas) y dispersión (desviaciones estándar, rangos), ya sea en la forma de variables continuas y como proporciones para las variables categóricas. Esta comparación es útil para describir la muestra de sujetos que entraron en el estudio. Pruebas de significancia (usualmente prueba de *t* y *chi* cuadrada) se reportan con valores de *p*. Cuando ocurren diferencias entre los grupos de tratamiento, no necesariamente se deben a una falla en la aleatorización, sino a un accidente. Altman[58] recomienda que se comparen basalmente los descriptores, por medio de una combinación de conocimiento clínico y sentido común. Si los grupos no están balanceados, debe llevarse a cabo un análisis no ajustado y otro ajustado por la variable que desbalanceó los grupos. Otra estrategia de análisis bajo esta circunstancia es hacer un análisis de doble diferencia, o ajustarlo por el desbalance.

El método estadístico usado para comparar dos medias en los diferentes grupos de tratamiento es la prueba *t* de Student. Esta prueba compara la media de dos grupos de variables continuas y expresa la probabilidad de que cualquier diferencia se deba al papel del azar (acepta la hipótesis nula) o que las diferencias son reales (rechaza la hipótesis nula). La prueba *t* tiene dos supuestos básicos: 1) que los datos en ambos grupos siguen una distribución normal, y 2) que para muestras no pareadas, la varianza para cada grupo es igual. Cuando estos supuestos no se cumplen, se recomienda el uso de pruebas no paramétricas. Cuando la variable tiene valores binarios (p. ej., curación o muerte) se puede calcular la proporción del evento para cada grupo de tratamiento. El estadístico más común es la prueba de *chi* cuadrada (también $ji^2$ o $X^2$). La $chi^2$ determina el número esperado del evento en ambos brazos de los grupos de tratamiento y los compara con los eventos observados y se expresa de la siguiente manera: $CHI^2=\Sigma(O–E)^2/E$, donde *O* representa las frecuencias observadas y *E* representa las frecuencias esperadas en cada celda de una tabla de contingencias de $2 \times 2$. Esta prueba puede usarse para comparar más de dos proporciones (cuadro IX).

## Análisis de supervivencia

Es el método estándar para analizar el tiempo que transcurre entre un evento inicial (que determina la inclusión del individuo en el estudio) y un evento final (genéricamente lla-

## Cuadro IX Ejemplo de comparación de dos proporciones

Resultados de un ensayo clínico aleatorizado para medir el efecto de la información sobre la elección de métodos de planificación familiar y riesgo de enfermedades sexualmente transmitidas en mujeres con infección cervical. (Tomado de: Lazcano-Ponce EC, Sloan NL, Winikoff B, Langer A, Coggins C, Heimburger A *et al.* The power of information and contraceptive choice in a family planning setting in Mexico. Sex Transm Inf 2000;76:277–281.)

Evento: Elección de dispositivo intrauterino (DIU) después de la intervención (información).

Grupo 1 (control): 914 eventos (88.5%) de un total de 1 033 pacientes.

Grupo 2 (intervención): 625 eventos (58.2%) de un total de 1 074 pacientes.

$$z = \frac{\hat{p}1 - \hat{p}2}{\sqrt{\hat{p}c(1-\hat{p}c)(1/n1 + 1/n2)}}$$

$H_o$: $\Pi 1 = \Pi 2$

$H_a$: $\Pi 1 \neq \Pi 2$

Donde $n_1$ y $n_2$ son el tamaño de muestra (1 033 y 1 074), $x_1$ y $x_2$ el número de eventos en los dos grupos (914 y 625), respectivamente, y

$$\hat{p}1 = x1/n1 (= 914/1033) = 0.8848$$

$$\hat{p}2 = x2/n2 (= 625/1074) = 0.5819$$

$$\hat{p}c = \frac{x1 + x2}{n1 + n2} (= [914 + 625)/1033 + 1074)] = 0.7304$$

$$z = \frac{0.8848 - 0.5819}{\sqrt{0.7304(1 - 0.7304)(1/1033 + 1/1074)}} = 15.70$$

Entonces, la hipótesis nula puede estar reflejada ($H_o$: $\Pi_1 = \Pi_2$) en el nivel de significancia de $\alpha$ si: $|z| \geq z\ (\alpha/2)$, donde $z\ (\alpha/2$ es el percentil superior $(\alpha/2)$ de la distribución normal estándar. En este caso la $z = 15.70$ lo cual se puede equiparar a una probabilidad de $p < 0.001$.

La diferencia estimada es 0.3029, por lo que existen diferencias entre el grupo control y el de intervención.

Los resultados sugieren que menos mujeres en el grupo de intervención seleccionaron el dispositivo intrauterino comparado con mujeres en el grupo control, esto es, la intervención redujo la selección del DIU, especialmente para mujeres con infecciones cervicales en quienes el DIU está contraindicado.

mado falla) que ocurre cuando el individuo presenta la característica para terminar el estudio (muerte, alta de la enfermedad, etc.). Se trata de un método flexible y puede medir múltiples eventos por sujeto, así como tomar en consideración en el análisis la información completa y parcial cuando así lo determina la ocurrencia de censuras (administrativas por término del estudio o logro del número de sujetos requeridos, y no administrativas por

pérdidas en el seguimiento). El estimador de Kaplan-Meier[59] toma en cuenta estas censuras al estimar las tasas de supervivencia entre los grupos de tratamiento, asumiendo que la censura es no informativa (esto es, que las censuras no administrativas no están relacionadas con la ocurrencia del evento en estudio). Para medir diferencias en las curvas de supervivencia entre los grupos de comparación, se toman en consideración los criterios de la prueba de Mantel y Haenszel (llamada también prueba de log rank).[60] Esta prueba pondera cada evento por igual y está cercanamente relacionada con las pruebas estadísticas del modelo de Cox,[61] también conocido como de riesgos proporcionales. Este modelo ofrece la ventaja de contabilizar las variables basales y pronósticas y su estimador (función de *hazard*) es la tasa instantánea de morir para aquellos sujetos que sobrevivieron hasta el tiempo *t*, y especifica cómo cambia la función de riesgo básica (individuos con nivel de covariables cero) respecto de aquellos con covariables distintas de cero. El cambio lo especifica el parámetro asociado a cada factor introducido al modelo y se interpreta como el cambio esperado en el cociente de riesgos entre un individuo en la población básica y uno fuera de ella, asumiendo que la razón de riesgos, para cualquier variable X, es constante a través del tiempo. Otro supuesto para cumplir con los modelos es que la curva de supervivencia de un grupo debe estar siempre por encima de la del otro grupo; en otras palabras, éstas no pueden cruzarse. El supuesto de proporcionalidad se evalúa mediante: a) las líneas del gráfico de las curvas de supervivencia (éstas no deben cruzarse), y b) las líneas del gráfico log-log: Ln[-Ln(S)] *vs.* tiempo para todos los grupos (las líneas deben ser aproximadamente paralelas). Si se violan los supuestos del modelo, entonces deben probarse modelos estratificados o es necesario buscar la interacción.

**Regresión**

El análisis de regresión mide la fuerza de asociación de una relación entre dos variables cuando una depende de la otra, bajo el supuesto de que esta relación es lineal. Si el análisis requiere que se estudien más variables se puede extender el modelo a un análisis de regresión lineal múltiple.

Cuando se requiere que el análisis utilice datos longitudinales, se pueden aplicar modelos de regresión mixtos, especialmente, porque pueden incorporar efectos fijos o aleatorios y permiten determinar diferencias en las estructuras de covarianza. Estos modelos son una extensión de los modelos de regresión lineal para variables continuas y se representan: $Y_i = X_i\beta + Z_i b_i + \epsilon_i$, donde $Y_i$ es el vector de respuesta para el sujeto pésimo sobre el periodo puntual del seguimiento, $X_i$ es la matriz del diseño usual para modelos fijos (tales como el tratamiento, variables basales, etc.) con su correspondiente coeficiente de regresión $\beta$, $Z_i$ es la matriz del diseño para efectos aleatorios (tal como un individuo tiende a desviarse de la media) con su correspondiente coeficiente $b_i$ y $\epsilon_i$, es el vector de los residuales. Una regresión aleatoria permite a cada sujeto tener su propia tendencia en el tiempo y el efecto del tratamiento puede causar diferencias en la pendiente sobre el tiempo. Cuando la variable resultado tiene estimaciones fuera de lo normal, pueden utilizarse modelos lineales generalizados y algunas de sus extensiones.[62]

Existen otros métodos de análisis que no han sido comentados en este capítulo y que se desarrollan en los capítulos de la sección de estadística; sin embargo, lo que hemos presentado hasta aquí nos da una idea de cómo ha evolucionado el análisis de los ensayos clínicos aleatorizados, hasta llegar a una alta sofisticación que no rebasa la experiencia del científico que utiliza métodos tradicionales efectivos plasmados en la planeación del diseño del estudio.

## ÉTICA DE LA INVESTIGACIÓN EN ENSAYOS CLÍNICOS ALEATORIZADOS

Proteger los derechos y el bienestar de los que participan en investigaciones científicas constituye el propósito actual de la ética de la investigación. Lamentablemente, la historia ha señalado investigaciones que no consideraron a los participantes como personas sino como objetos de estudio. Por esta razón, la Declaración de Helsinki,[63] de 1964, estableció uno de los primeros antecedentes en principios éticos de investigaciones en seres humanos. A este respecto, el artículo 5º expresó su preocupación por el bienestar de los seres humanos, el cual debe tener [siempre] supremacía sobre los intereses de la ciencia y de la sociedad. El cuadro X muestra de manera cronológica el desarrollo histórico de las guías internacionales para la investigación con seres humanos.



### Consentimiento informado y genuino

Un ensayo clínico exitoso depende, entre muchos factores, del proceso de obtención del consentimiento informado (CI). La tensión generada entre la necesidad de seleccionar a los participantes y la obligación de ofrecer ciertos tipos de protección ha hecho del CI un persistente reto ético, ya que (en los países en desarrollo, por dar un ejemplo) las costumbres, las tradiciones y las concepciones que tienen acerca de la salud y la enfermedad pueden variar significativamente, además del posible desconocimiento relacionado con la investigación biomédica. De acuerdo con diversos acuerdos internacionales, el CI consiste en la aceptación para participar en un trabajo de investigación, de un individuo competente que ha recibido la información necesaria, la ha comprendido adecuadamente y, después de considerarla, ha decidido participar en el estudio, sin haber sido coercionado, intimidado ni sometido a influencias o incentivos indebidos.[64] Idealmente, esta decisión forma parte de un proceso que incluye la discusión del estudio con el investigador principal (IP) así como con otros colaboradores, y la garantía de que la información fue comprendida por el participante para que, finalmente, el paciente o su representante legal otorgue la firma en el documento de CI (cuadro XI).

### Comisiones de Ética [CE]

Universalmente se acepta que no debe hacerse investigación médica en personas sin seguir los postulados y lineamientos nacional e internacionalmente aprobados por la ley. En México, la Ley General de Salud, en materia de investigación, señala en su artículo 99 que toda institución de salud en donde se realice investigación deberá tener una Comisión de ética (CE) si la realiza con seres humanos.[65] De acuerdo con las Guías operacionales para CE que evalúan investigación biomédica, propuestas por la Organización Mundial de la Salud (OMS), el propósito de una CE es contribuir a salvaguardar la dignidad, los derechos, la seguridad y bienestar de todos los y las participantes actuales y potenciales de la investigación.[66]

Las CE existen para asegurar que la investigación propuesta responda a las necesidades de salud de la población, sin exponer a los participantes a riesgos inaceptables e innecesarios, y que los participantes potenciales tengan la garantía de ser completamente informados y, por lo tanto, la capacidad para evaluar las consecuencias previstas de su participación y decidan su ingreso al estudio, mediante un consentimiento genuino.[67]

### Conferencia internacional de armonización y buenas prácticas clínicas

Las *buenas prácticas clínicas* (BPC) son un estándar para el diseño, conducción, desarrollo, monitoreo, auditoría, registro, análisis y repor-

**Cuadro X Guías internacionales para la investigación con seres humanos**

| *Año* | *Organización* | *Título del documento* |
|---|---|---|
| 1947 | Crímenes de guerra en el Tribunal de Nuremberg | Código de Nuremberg |
| 1948 | Asamblea General de las Naciones Unidas | Declaración Universal de los Derechos Humanos |
| 1964 | Asociación Médica Mundial (AMM) | Declaración de Helsinki (1) |
| 1975 | AMM | Declaración de Helsinki (2) Tokio |
| 1983 | AMM | Declaración de Helsinki (3) Venecia |
| 1989 | AMM | Declaración de Helsinki (4) Hong Kong |
| 1991 | CIOMS-OMS | Guías internacionales para la revisión ética de los estudios epidemiológicos |
| 1993 | CIOMS-OMS | Pautas éticas internacionales para la investigación biomédica en seres humanos |
| 1995 | OMS | Buenas prácticas clínicas [BPC] en ensayos de productos farmacéuticos |
| 1996 | AMM | Declaración de Helsinki (5) Sudáfrica |
| 1996 | Conferencia Internacional de Armonización (CIA) | Guía de armonización tripartita-BPC. |
| 1997 | Consejo de Europa | Convención para la prevención de los derechos y dignidad de los seres humanos respecto de la aplicación de la medicina y la biología |
| 1997 | UNESCO | Declaración Universal sobre el Genoma Humano y los Derechos Humanos |
| 2000 | Unión Europea | Declaración de los Derechos Fundamentales para la Unión Europea |
| 2000 | ONUSIDA | Consideraciones éticas para las investigaciones de vacunas preventivas |
| 2000 | OMS | Guías operacionales para Comités de Ética que evalúan investigación biomédica |
| 2000 | AMM | Declaración de Helsinki (6) Edimburgo |
| 2001 | Parlamento Europeo y Consejo de la Unión Europea | Directiva 2001/20/EC. Leyes, regulaciones y provisiones administrativas de los Estados Miembro, relacionadas con las *buenas prácticas clínicas* en la conducción de ensayos clínicos de productos médicos empleados para el uso en humanos |

**Cuadro XI Consentimiento informado y genuino. Requerimientos internacionales para la investigación con seres humanos**

| *Declaración de Helsinki (2000)*[a] |
|---|
| 1. Objetivos del estudio y los métodos que serán empleados. |
| 2. Fuentes de financiamiento y posibles conflictos de interés. |
| 3. Filiaciones institucionales del investigador. |
| 4. Beneficios anticipados y riesgos potenciales, así como el seguimiento del estudio. |
| 5. Incomodidad, dolor, riesgo, o inconveniente previsibles. |
| 6. Derecho a abstenerse de participar o retirarse del estudio sin ninguna represalia. |
| 7. Verificación de parte del investigador de que la información es comprendida por el participante. |
| 8. Obtención del consentimiento preferiblemente de forma escrita. De no ser por escrito, el investigador deberá documentarlo y presentarlo a la Comisión de Ética. |
| *CIOMS (2002)*[b, c] |
| 1. Disposición de cualquier producto o intervención de efectividad y seguridad comprobadas por la investigación; cuándo y cómo estará disponible, y si se espera que paguen por él. |
| 2. Medidas para asegurar el respeto a la privacidad de los sujetos y a la confidencialidad de los registros en los que se identifica al sujeto. |
| 3. Normas relativas al uso de resultados de pruebas genéticas e información genética familiar, y las precauciones tomadas para prevenir la revelación de los resultados, sin el consentimiento del sujeto. |
| 4. Especificación sobre si el investigador está actuando únicamente como investigador o como investigador y médico del sujeto. |
| 5. Definición del grado de responsabilidad del investigador encargado de proporcionar servicios médicos al participante. |
| 6. Especificación sobre si se compensará al sujeto, a su familia o a sus dependientes en caso de discapacidad o muerte como resultado de estos daños y por medio de qué mecanismo y organización se hará (o, cuando corresponda, que no habrá lugar a compensación). |
| En el caso de ensayos clínicos que son financiados externamente. Obtención del consentimiento informado: obligaciones de patrocinadores e investigadores [d] |
| 1. Evitar el engaño, influencia indebida o intimidación. |
| 2. Solicitar el consentimiento sólo después de comprobar que el sujeto potencial para el estudio ha comprendido adecuadamente los hechos relevantes y las consecuencias de su participación, y ha tenido suficiente oportunidad de considerarlo. |
| 3. Obtener por escrito un formulario de consentimiento informado de cada sujeto potencial. |
| 4. Renovar el consentimiento informado de cada sujeto, si se producen cambios significativos en las condiciones o procedimientos. |
| Lineamientos para las *buenas prácticas clínicas* (BPC)[c,e] |
| 1. Explicar al participante la probabilidad de una asignación aleatoria a cada tratamiento. |
| 2. Informarle que él o su representante legal aceptado tendrán acceso directo al monitor, auditor, a la Comisión de Ética (CE) y a las autoridades regulatorias de los registros médicos originales del sujeto para la verificación de los procedimientos y/o datos del estudio clínico, sin violar la confidencialidad del sujeto hasta donde lo permitan las leyes y regulaciones aplicables, al firmar una carta de consentimiento informado. |

**Cuadro XI** ***Continuación***

3. Proporcionar el número de sujetos aproximado involucrados en el estudio.

4. En situaciones de emergencia, cuando no sea posible obtener el CI previamente, deberá pedirse el CI del representante legal aceptado, si lo hubiere. Si éste no está disponible, la inclusión del sujeto requerirá de las medidas descritas en el protocolo y/o en otra parte con la aprobación de la CE.

a. World Medical Association. (1996, 2000). Declaration of Helsinki. Edinburgh, Scotland. Paragraph No. 22.
b. Council for International Organizations of Medical Sciences, 1993; 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland. Guideline No. 5.
c. Además de las incluidas en la Declaración de Helsinki.
d. Council for International Organizations of Medical Sciences, 1993; 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland. Guideline No. 6.
e. Buenas prácticas clínicas. Lineamientos de la Conferencia Internacional de Armonización [CIARM] sobre requerimientos técnicos para el registro de productos farmacéuticos para uso en humanos. Guía tripartita Estados Unidos de América, Japón y la Comunidad Europea [EMEA]. Cabe aclarar que los lineamientos de las BPC están directamente relacionados con investigaciones de productos farmacéuticos y, por lo tanto, la mayoría de ellos se traslapan con CIOMS y la Declaración de Helsinki; sin embargo, algunos otros se corresponden con esta modalidad de investigación.

te de estudios de investigación clínica en los que participan seres humanos como sujetos de estudio. El cumplimiento de las BPC asegura dos aspectos fundamentales en los que se pone a prueba un medicamento o un equipo de diagnóstico previo a su comercialización: a) la protección de los derechos, la seguridad y el bienestar de los sujetos de estudio, y b) la credibilidad de los datos generados, necesaria para el registro de una innovación terapéutica en cualquier parte del mundo.

Los lineamientos para la conducción de Estudios Clínicos, conocidos como ICH (por sus siglas en inglés) constituyen el estándar bajo el cual se rigen los países originalmente incluidos para llevar a cabo investigación en seres humanos y, por lo tanto, se sugiere que cualquier grupo de investigación que desee realizar estudios clínicos y aportar datos que sean válidos en Estados Unidos, la Unión Europea y el Japón, debe acogerse a estos lineamientos.[68] Las recomendaciones de ICH se dividen en cuatro categorías con temas específicos en cada una: Q, para calidad; S, para seguridad; E, para eficacia, y M, para temas multidisciplinarios. En la Sección ICH de eficacia se encuentran todos los temas relacionados con aspectos clínicos.



### Buenas prácticas clínicas

Los lineamientos de BCP son el estándar que se sigue durante el desarrollo de proyectos de investigación clínica. A este respecto, se encuentran claramente especificadas las responsabilidades y funciones del investigador y el patrocinador; los contenidos básicos del protocolo de investigación del ensayo clínico; el manual del investigador, y los documentos esenciales para la conducción de un estudio clínico. Aquí se cubren todos los aspectos de preparación, supervisión, notificación y archivo de estudios clínicos, y se incluyen los lineamientos de BPC de la Unión Europea, Japón, Estados Unidos, Australia, Canadá, Países Nórdicos y la OMS, que deben ponerse en práctica cuando se generan datos clínicos que pretendan someterse a autoridades regulatorias o en cualquier proyecto de investigación clínica que pueda tener un impacto en la seguridad y bienestar de los sujetos participantes.

### Responsabilidades del investigador

El investigador principal es, por definición, el encargado de vigilar el cumplimiento de los lineamientos de ICH, el respeto a las leyes locales y el apego al protocolo. Debe organizar un equipo de trabajo que le ayude a cumplir las expectativas en cuanto al número de pacientes reclutados, tiempos para llevar a cabo el proyecto y calidad de los datos generados. En el cuadro XII, se describen las responsabilidades básicas del IP.

### Responsabilidades del patrocinador

El patrocinador es responsable de implementar y mantener sistemas para un aseguramiento de la calidad y control de calidad con procedimientos estándar de operación, descritos para asegurar que se conduzcan los estudios y se generen, documenten y notifiquen los datos, en cumplimiento con el protocolo, las BPC y los requerimientos regulatorios que apliquen (cuadro XIII). Todos los acuerdos que lleve a cabo el patrocinador con el investigador o la institución, y/o con cualquier otra parte involucrada en el estudio clínico, deberán hacerse por escrito. El programa de control de calidad implementado por el patrocinador debe incluir actividades tanto a nivel interno como externo, de manera que se encuentren correctamente documentados todos los procedimientos que tienen que ver con el estudio. Dicho programa debe incluir visitas regulares de supervisión a los centros de investigación, detección oportuna de errores en la conducción del estudio y corrección inmediata de los mismos, con estrategias que aseguren su prevención a futuro, que incluyan entrenamiento y reentrenamiento, en caso necesario.

### La supervisión como una de las responsabilidades del patrocinador

Uno de los elementos más importantes de la investigación clínica es la supervisión y doble verificación de los datos generados. El trabajo del *monitor clínico* tiene el objetivo primordial de cuidar que estén protegidos los derechos y el bienestar de los seres humanos, al mismo



**Cuadro XII Actividades y responsabilidades del investigador**

1. Vigilar en todo momento el cuidado de los pacientes, asegurándose de que cada uno de ellos ha firmado su consentimiento informado antes de incluirlo en el estudio:
   a. las decisiones médicas las debe tomar el personal médico calificado;
   b. los pacientes deben recibir tratamiento médico adecuado en caso de eventos adversos;
   c. el código ciego sólo debe romperse de acuerdo con el protocolo;
   d. en caso de que se presentara un evento adverso, el código ciego sólo debe abrirse si la decisión terapéutica lo amerita, y
   e. debe documentarse y explicársele de manera inmediata al patrocinador cualquier situación relacionada con los pacientes, que pudiera influir en su decisión para seguir participando en el estudio.
2. Supervisar que el protocolo se cumpla al pie de la letra:
   a. el estudio debe conducirse, exactamente, de acuerdo con el protocolo aprobado;
   b. no debe haber ninguna desviación o cambio a lo que dice el protocolo, sin el acuerdo del patrocinador y la aprobación de la Comisión, excepto cuando haya riesgo para los sujetos participantes, y
   c. debe explicarse y documentarse cualquier desviación al protocolo.

**Cuadro XII** ***Continuación***

3. Mantener una comunicación constante con la Comisión de Ética, y procurar su aprobación en caso necesario:
   a. hay que tener la aprobación por escrito antes de comenzar el estudio y de manera específica de cada uno de los siguientes documentos: protocolo, consentimiento informado, procedimientos de reclutamiento y de cualquier material que se le entregue al paciente;
   b. durante el estudio, la Comisión debe recibir todos los documentos sujetos a revisión o cualquier cambio o desviación en el protocolo aprobado, y
   c. deben someterse a la Comisión de Ética los resúmenes por escrito del estado del estudio, por lo menos una vez al año (o con mayor frecuencia, en caso necesario).

4. Vigilar que el medicamento de estudio se utilice de manera adecuada, y asegurar su correcto almacenamiento e inventario:
   a. el conteo del medicamento es responsabilidad del investigador. Puede asignarse a una persona calificada, pero bajo su supervisión;
   b. deben mantenerse registros detallados y con reconciliaciones cuidadosas de medicamento recibido, en inventario, usado y regresado;
   c. el medicamento debe almacenarse de acuerdo con las especificaciones del patrocinador (bajo llave);
   d. debe asegurarse que el medicamento sólo se utiliza de acuerdo con el protocolo, y
   e. debe explicárseles a los pacientes el uso correcto del medicamento, y verificar su apego de manera regular.

5. Elaborar o supervisar la elaboración de los reportes necesarios durante el transcurso del estudio:
   a. deben entregarse por escrito, de manera oportuna, todos los informes que requieran el patrocinador, el Comité o la Institución;
   b. los datos deben reportarse a tiempo, de manera correcta, completa y legible;
   c. los datos reportados en el cuaderno de trabajo deben ser consistentes con los documentos fuente, y debe explicarse cualquier discrepancia;
   d. cualquier cambio o corrección en un cuaderno de trabajo o en un documento fuente debe fecharse y establecerse de inicio, por medio de una línea en la entrada original (nunca debe borrarse o tacharse completamente);
   e. los eventos adversos serios deben reportarse de manera inmediata al patrocinador, seguidos de reportes escritos detallados, dirigidos a quien corresponda, y
   f. los eventos especiales (por protocolo) deben reportarse como lo especifique el patrocinador (eventos adversos serios).

6. Mantener los documentos del estudio en la forma y por el tiempo indicados en cada caso:
   a. todos los documentos del estudio deben mantenerse y estar disponibles cuando sean requeridos por el personal autorizado, y
   b. los documentos esenciales deben retenerse el tiempo que sea necesario. Las recomendaciones de ICH dicen textualmente que los documentos deben retenerse "por lo menos dos años después de la última aprobación de una aplicación de mercadotecnia en una región de ICH, hasta que no haya aplicaciones pendientes o planeadas, o hasta que hubieran pasado, por lo menos, dos años a partir de la suspensión del desarrollo clínico del producto". La Ley General de Salud en México exige que los documentos se mantengan por un mínimo de 15 años.

**Cuadro XIII Actividades y responsabilidades del patrocinador**

*(Sección ICH-E6, Capítulo 5)*

1. Aseguramiento y control de calidad.
2. Organizaciones de investigación por contrato (Clinical Research Organizations, CRO).
3. Experiencia y conocimiento médico.
4. Diseño del estudio.
5. Organización gerencial del estudio, manejo de los datos y mantenimiento de registros.
6. Selección de investigadores.
7. Distribución de responsabilidades.
8. Compensación a los sujetos y a los investigadores.
9. Financiamiento.
10. Notificación a las autoridades regulatorias.
11. Confirmación de revisión y aprobación del Comité de Ética.
12. Información acerca del producto de investigación.
13. Manufactura, empaquetamiento, etiquetado y codificación del medicamento.
14. Abastecimiento y manejo del producto de investigación.
15. Acceso a registros.
16. Información sobre seguridad.
17. Reporte de reacciones adversas medicamentosas.
18. Supervisión.
19. Auditoría.
20. Vigilancia de adherencia al protocolo (falta de cumplimiento).
21. Terminación prematura o suspensión del estudio.
22. Reportes generales o finales del estudio.
23. Ensayos multicéntricos.



tiempo que vigila que estén completos los datos generados por el estudio, que sean precisos y puedan verificarse en los documentos fuente de acuerdo con el protocolo, las BPC y los requerimientos regulatorios aplicables.

### Manejo ético de la información

La información es, en realidad, el producto final que se persigue con la conducción del estudio clínico. Los datos generados durante el estudio deben tener la calidad que se busca, y estar disponibles en los tiempos requeridos para su análisis. Desde su inicio, todo estudio clínico tiene un plan de manejo de datos que implica tiempos establecidos previamente para su procesamiento, que incluye captura, limpieza y análisis de la información.

Cuando se utilicen sistemas de manejo electrónico de datos, el patrocinador deberá asegurar que el sistema esté diseñado de manera tal que permita efectuar cambios en los datos; que no se borren los registros originales, y tenga un respaldo y una documentación adecuados. Además, debe haber un sistema de seguridad que impida el acceso no autorizado a los datos.

## CONCLUSIONES

Los ensayos clínicos aleatorizados son los estudios más cercanos al método experimental, por lo que forman el paradigma de la investigación epidemiológica. Representan el diseño con el nivel más alto de causalidad, donde el investigador tiene control sobre la exposición; son prospectivos; tienen la ventaja de que, en teoría, pueden evitarse los sesgos, y tienen un elevado nivel de comparabilidad de poblaciones (confusión) así como de información. Su práctica, sin embargo, es muy compleja, y en países en desarrollo se le debe dar prioridad, además de a los aspectos metodológicos, a la garantía de asegurar los derechos de los pacientes.

## REFERENCIAS


1. Packard FR. Life and times of Ambroise Paré, 1510-1590. Paul B. Hoeber, Nueva York, 1921.
2. Good P. A manager's guide to the design and conduct of clinical trials. Wiley-Liss Inc., 2002: 47-64.
3. Greeno C. Major alternatives to the classic experimental design. Fam Process 2002;41:733-736.
4. Oxman MN, Levin MJ, Johnson GR, Schmader KE, Straus SE, Gelb LD, Arbeit RD, *et al.,* A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. N Engl J Med. 2005 Jun 2;352(22):2271-2284.
5. Barbui C, Violante A, Garattini S. Does placebo help establish equivalent in trials of new antidepressants? Eur Psychiat 2000;15:268-273.
6. Rivera JA, Sotres-Álvarez D, Habicht JP, Shamah T, Villalpando S. Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children: a randomized effectiveness study. JAMA 2004;291:2563-2570.
7. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol 2003;3:1-5.
8. Menendez C, Kahigwa E, Hirt R, Vounatsou P, Aponte J, Font, *et al.* Randomised placebo-controlled trial of iron supplementation and malaria chemoprophylaxis for prevention of severe anaemia and malaria in Tanzanian infants. The Lancet 1997;350:844-850.
9. Mundorf TK, Ogawa T, Naka H, Novack GD, Crockett RS; US Istalol Study Group. A 12-month, multicenter, randomized, double-masked, parallel-group comparison of timolol-LA once daily and timolol maleate ophthalmic solution twice daily in the treatment of adults with glaucoma or ocular hypertension. Clin Ther 2004;26:541-551.
10. Breitburd F, *et al.,* Immunization with viruslike particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. J Virol 1995;69:3959–3963.
11. Villa LL, Costa , *et al.* Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncol 2005;6(5):271-278.
12. Stanley MA. Human papillomavirus vaccines. Rev Med Virol 2006;16(3):139-149.
13. Traversa G, Bignami G. Ethics problems in phase IV of drugs studies. Ann Ist Super Sanita 1998;34:203-208
14. Linden M. Differences in adverse drug reactions in phase III and phase IV of the drug evaluation process. Psychopharmacol Bull. 1993;29: 51-56
15. Moher D, Schultz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-groups randomised trials. JAMA 2001;357:1191-1194.
16. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health ser-

vices. The Evidence-Based Medicine Working Group. JAMA 1996;275:554-558.
17. Begg CB, Cho MK, Eastwood S, Horton R, Moher D, Olkin I, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-639.
18. Schulz K, Grimes D. Blinding in randomised trials: hiding who got what. The Lancet 2002;359:696-700.
19. Emanuel EJ., Miller FG. The Ethics of placebo-controlled trials. A middle ground. N Engl J Med 2001;345:915-918.
20. Schafer A. The ethics of the randomized clinical trials. N Engl J Med 1982;307:719-724.
21. Freedman B. Placebo-controlled trials and the logic of clinical purpose. IRB 1990;12:1-6.
22. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001;344:1594-1602.
23. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? J Clin Epidemiol 1997;50:1311-1318.
24. Beecher HK. The powerful placebo. J Am Med Assoc 1955 Dec 24;159:1602-1606.
25. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Engl J Med 1994;331:394-398.
26. Karrison TG, Huo D, Chappell R. A group sequential, response-adaptive design for randomized clinical trials. Control Clin Trials 2003;24:506-522.
27. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:1-173.
28. Krause MS, Howard KI. What random assignment does and does not do. J Clin Psychol 2003 Jul;59:751-766.
29. Berger VW, Bears JD. When can a clinical trial be called "randomized"? Vaccine 2003 17;21:468-472.
30. Soares I, Carneiro AV. Intention-to-treat analysis in clinical trials: principles and practical importance. Rev Port Cardiol 2002;21:1191-1198.
31. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. A review. Control Clin Trials 2002;23:662-674.
32. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet. 2002;359:614-618.
33. Slack MK, Draugalis JR. Establishing the internal and external validity of experimental studies. Am J Health Syst Pharm 2001;58:2173-2181.
34. Altman DG, Schulz KF. Statistics notes: Concealing treatment allocation in randomised trials. BMJ 2001;323:446-447.
35. Cotton PB. Randomization is not the (only) answer: a plea for structured objective evaluation of endoscopic therapy. Endoscopy 2000;32:402-405.
36. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999;319:670-674.
37. Zelen M, Lee SJ. Models and the early detection of disease: methodological considerations. Cancer Treat Res 2002;113:1-18.
38. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd. ed. Nueva York: Springer-Verlag, 1998.
39. Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. Epidemiol Rev 2002;24:26-38.
40. Green SB. Design of randomized trials. Epidemiol Rev 2002;24:4-11.
41. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. The Anturane Reinfarction Trial. N Engl J Med 1978;298:289-295.
42. Sulfinpyrazone in the prevention of sudden death after myocardial infarction. The Anturane Reinfarction Trial Research Group. N Engl J Med 1980;302:250-256.
43. Temple R, Pledger GW. The FDA's critique of the anturane reinfarction trial. N Engl J Med 1980;303:1488-1492.
44. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 1980;303:1038-1041.
45. DeMets DL. Statistical issues in interpreting clinical trials. J Intern Med 2004;255:529-537.
46. Wang D, Bakhai A, Maffulli N. A primer for statistical analysis of clinical trials. Arthroscopy 2003;19:874-881.

47. Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. Control Clin Trials 1997;18:204-221.
48. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979:65-70.
49. Westfall PH, Young SS. P-value adjustments for multiple tests in multivariate binomial models. J Am Stat Assoc 1989;84:780-786.
50. Mantel N. Assessing laboratory evidence for neoplastic activity. Biometrics 1980;36(3):381-399.
51. Tukey JW, Ciminera JL, Heyse JF. Testing the statistical certainty of a response to increasing doses of a drug. Biometrics 1985;41(1):295-301.
52. O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics 1984;40:1079-1087.
53. Henderson WG, Fisher SG, Cohen N, Waltzman S, Weber L. Use of principal components analysis to develop a composite score as a primary outcome variable in a clinical trial. The VA Cooperative Study Group on Cochlear Implantation. Control Clin Trials 1990;11(3): 199-214.
54. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229-236.
55. Dickstein K, Kjekshus J. Effects of losartan and captopril on mortality and morbidity in high-risk patients after acute myocardial infarction: the OPTIMAAL randomised trial. Optimal Trial in Myocardial Infarction with Angiotensin II Antagonist Losartan. Lancet 2002;360: 752-760.
56. Perrone F, Di Maio M, De Maio E, Maione P, Ottaiano A, Pensabene M, *et al*. Statistical design in phase II clinical trials and its application in breast cancer. Lancet Oncol 2003;4:305-311.
57. Sylvester R, Van Glabbeke M, Collette L, Suciu S, Baron B, Legrand C, *et al*. Statistical methodology of phase III cancer clinical trials: advances and future perspectives. Eur J Cancer 2002;38 Suppl 4:S162-S168.
58. Altman DG. Comparability of randomized groups. Statistician 1985:125-136.
59. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958:457-481.
60. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966:163-170.
61. Cox DR. Regression models and life tables. J R Stat Soc, series B, 1972;34:187-220
62. Breslow NR, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1985;88:9-25.
63. World Medical Association. (1996, 2000). Declaration of Helsinki. Edinburgo. Disponible en http://www.wma.net/e/policy/17-ce.html.
64. Council for International Organizations of Medical Sciences, 1993; 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland. 2002 Revisión disponible en http://www.cioms.ch.
65. Reglamento de la Ley General de Salud en Materia de Investigación para la Salud. [1986]. Disponible en http://www.salud.gob.mx/unidades/cdi/nom/compi/rlgsmis.html
66. Gilbert C, Fulford KW, Parker C. Diversity in the practice of district ethics committees. B M J 1989;299:1437-1439.
67. Torgerson DJ., Dumville JC. Research governance also delays research. BMJ 2004;328:710.
68. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) adopts Consolidated Guideline on Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use. Int Dig Health Legis 1997;48:231-234.

# Ensayos clínicos aleatorizados — Ejercicios

Las preguntas de este ejercicio se basan en el siguiente artículo:

## REFERENCIA

Kershenobich D, Vargas F, García-Tsao G, Pérez Tamayo R, Gent M, Roffind M. Colchicine in the treatment of cirrhosis of the liver. N Eng J Med 1988;318:1709-1713.

## RESUMEN

Generalmente, el tratamiento de la cirrosis hepática es sintomático. Sin embargo, la colchicina, un inhibidor de la síntesis de colágena, puede ser benéfica en el tratamiento de cirrosis hepática. Para evaluar el efecto del uso de colchicina en la sobrevida de pacientes con cirrosis hepática, los autores desarrollaron un ensayo clínico aleatorizado, doblemente ciego, controlado con placebo.

## PREGUNTAS

1. ¿Cuál de las tres características mencionadas en el resumen considera usted que es la que distingue a los estudios observacionales de los experimentales? Comente su respuesta.
2. ¿Qué busca el investigador al asignar de manera aleatoria a los sujetos a los diferentes grupos de comparación? Comente.
3. ¿Cuál considera usted que sería el principal criterio de inclusión para evaluar la sobrevida de pacientes con cirrosis hepática?
4. Después de dar su consentimiento informado, cien sujetos fueron elegibles para el estudio. ¿Cuántos grupos de comparación establecería usted? Comente.
5. Los autores realizaron un análisis preliminar en donde evaluaron la eficacia del tratamiento, con base en la mortalidad por cualquier causa. En este análisis, los autores incluyeron ocho sujetos que fueron diagnosticados incorrectamente como cirróticos y todos aquellos que murieron independientemente de haber tomado correctamente el medicamento (21 muertes en el grupo de intervención y 28 en el grupo control). Usaron el método de "intención de tratar". Explique en qué consiste esta metodología y explique qué quisieron evitar los investigadores al utilizarla.
6. ¿Cree usted que este tipo de análisis sobre o subestimó la eficacia de la colchicina?
7. El documento denominado consentimiento informado debe establecer, entre otras cosas, que la participación de los sujetos de investigación es absolutamente voluntaria y que tienen el derecho de retirarse en cualquier momento.
   a) verdadero
   b) falso
8. Los autores reportaron 21 muertes en el grupo que recibió colchicina (n = 54) y 28 en el grupo que recibió placebo (n = 46). Estime: i) el riesgo de morir en ambos grupos de comparación; ii) el riesgo relativo; iii) la reducción del riesgo relativo; iv) la reducción del

riesgo absoluto, y v) el número de pacientes necesario a tratar con colchicina para prevenir la cirrosis hepática. Indique el procedimiento e interprete sus resultados.

9. Los autores reportaron que la sobrevida en el grupo que recibió colchicina fue de 11 años, mientras que la del grupo control fue de 3.5 años. La diferencia fue estadísticamente significativa. ¿A quién considera usted que podría extrapolar los resultados? Comente.

## RESPUESTAS

1. La aleatorización (asignación aleatoria de la exposición) es la principal característica que distingue los estudios observacionales de los experimentales. En éstos es el investigador quien asigna, de manera aleatoria, a cada uno de los individuos que participan en el estudio al grupo de intervención o al grupo control. En contraste, en los estudios observacionales, los individuos ya están expuestos o no a la característica de interés.
2. La asignación aleatoria sirve para que aquellos factores que podrían tener un efecto sobre la variable de interés y actuar como confusores se distribuyan de manera comparable en los diferentes grupos de comparación. Esto es, se controla por todas las características de los individuos, aun por las que no podemos observar o medir.
3. Tener el diagnóstico de cirrosis hepática. Los autores consideraron como criterios de inclusión tener un diagnóstico definitivo de cirrosis hepática y tener, por lo menos, 18 años de edad.
4. Dos grupos. Los sujetos incluidos en el estudio fueron asignados aleatoriamente al grupo de intervención, que consistió en tomar colchicina, y al grupo control, que consistió en tomar placebo.
5. El método de análisis por "intención de tratamiento" consiste en incluir en el análisis a todos los pacientes tal y como fueron asignados a los grupos de comparación. Se recomienda incluir en el análisis a todos los sujetos respetando la asignación aleatoria, de tal manera que aquellos sujetos cuyo pronóstico es diferente, permanezcan igualmente distribuidos en los grupos de comparación. El propósito de su realización es evitar sesgos de selección por pérdidas en el seguimiento.
6. La subestimó.
7. Verdadero.
8. Procedimiento y resultados:
   Riesgo de morir en el grupo de colchicina $(21/54) \times 100 = 38.9\%$.
   Riesgo de morir en el grupo control $(28/46) \times 100 = 60.9\%$.
   Riesgo relativo $38.9/60.9 = 0.64$.
   Reducción del riesgo relativo $(1 - 0.64) = 0.36$.
   Reducción del riesgo absoluto $60.9 - 38.9 = 22$.
   Número de pacientes necesarios para prevenir un probable evento $1/(0.61 - 0.39) = 4.5$.

   *Interpretación: el riesgo de morir en aquellos que recibieron colchicina fue de 38.9%; sin embargo, el riesgo de morir en el grupo que usó placebo fue mayor (60.9%).*

   *El riesgo de morir de los que recibieron colchicina fue 0.64 veces el riesgo de los que recibieron placebo.*

*Se observó una disminución del riesgo de 36% en aquellos sujetos que recibieron colchicina, respecto de aquellos que no la recibieron.*

*El uso de colchicina reduciría el riesgo de morir de 60.9% a 38.9%; el 36% se refiere al exceso de riesgo de morir de cirrosis hepática que podríamos evitar, si los pacientes tomaran colchicina.*

*Se requiere tratar a 4.5 pacientes con cirrosis hepática para evitar una muerte.*

9. Asumiendo que no hubo sesgos, los resultados pueden extrapolarse a sujetos con las mismas características de los enfermos incluidos en este estudio. Aunque la sobrevida es un buen indicador de la eficacia del medicamento, es necesario repetir el estudio en más grupos de pacientes para evaluar si hay consistencia en los resultados y poder generalizar los resultados a un grupo mayor de sujetos.

# VI Estudios de cohorte

*Eduardo Lazcano Ponce*
*Esteve Fernández*
*Eduardo Salazar Martínez*
*Mauricio Hernández Ávila*

***Cohorte:*** Del latín *cohors, cohortis:* "séquito, agrupación". Entre los romanos, cuerpo de infantería que comúnmente constaba de 500 hombres, y era la décima parte de una legión. Por lo general, los veteranos ocupaban la primera y última fila de la cohorte. Puede provenir del verbo latino *cohortari*, "arengar", toda vez que la fuerza de la cohorte se ajustó generalmente al número de hombres que podían escuchar juntos la voz del jefe que les dirigía la palabra.

El diseño de estudios de cohorte es uno de los más usados en la investigación epidemiológica. En este tipo de diseño, la unidad de observación y análisis es el individuo, y su característica primordial radica en que los sujetos a investigar se eligen a partir de la presencia o ausencia de la exposición que interesa investigar y la ocurrencia del evento se puede determinar de manera prospectiva o retrospectiva. En su concepción más simple, el diseño de cohorte implica la selección de un grupo de personas con cierta característica, consideradas como expuestas al factor de estudio; la selección de un grupo similar, pero sin la exposición, considerado como no expuesto; el seguimiento de ambos grupos durante un periodo específico en el que se registra el evento de interés, y finalmente la comparación de la magnitud y frecuencia con que ocurre el efecto en ambos grupos.

Los estudios de cohorte prospectivos se asemejan a los ensayos clínicos aleatorizados, en el sentido de que los sujetos de estudio se siguen durante el curso de la exposición hasta la aparición del evento que interesa; pero, a diferencia del ensayo clínico aleatorizado, donde el investigador asigna intencionalmente la exposición, en los estudios de cohorte el investigador sólo observa a los sujetos después de ocurrida la exposición.

Anteriormente, los estudios de cohorte eran conocidos como a) longitudinales, porque los sujetos eran observados, por lo menos, en dos ocasiones; b) prospectivos, lo que implicaba que los sujetos eran reclutados antes de la ocurrencia del evento y estudiados para vigilar la ocurrencia del mismo; o c) de incidencia, porque generaban información con la cual era posible estimar esta medida epidemiológica. Actualmente, sin embargo, la expresión correcta para identificar este diseño es simplemente "estudios de cohorte".

Los estudios de cohorte se han utilizado de manera clásica para determinar la ocurrencia de un evento específico: la población en estudio se sigue a través del tiempo, y se compara la incidencia del evento en individuos expuestos con la de los no expuestos. El seguimiento continúa hasta que se presenta una de las siguientes condiciones:

a) se manifiesta el evento de estudio (de salud o enfermedad): cuando ocurre esta condición, el individuo deja de contribuir a la cohorte (ya que al manifestar el evento deja de estar en riesgo), pero puede reingresar si se trata de un evento recurrente, como puede ser un episodio de diarrea;
b) los sujetos del estudio mueren;
c) los sujetos declinan continuar su participación en el estudio o se pierde el contacto por alguna razón, o
d) el estudio termina.

## CLASIFICACIÓN DE LOS ESTUDIOS DE COHORTE

Dependiendo de la relación temporal entre el inicio del estudio y la ocurrencia del evento, los estudios de cohorte se han clasificado como: prospectivos, retrospectivos (o históricos) y mixtos (o ambispectivos). En el segundo tipo el investigador reconstruye la experiencia de la cohorte en el tiempo, por lo que dependen de la disponibilidad de registros para establecer y medir exposición y el resultado. Una aplicación frecuente de una cohorte histórica son los estudios de exposición ocupacional, donde se reconstruye la exposición y frecuencia de eventos en un grupo ocupacional o empresa durante un periodo determinado. La validez de los estudios retrospectivos depende, en gran medida, de la calidad de los registros utilizados. En contraste, en las cohortes prospectivas, es el investigador quien documenta la ocurrencia del evento, por lo que la exposición se valora de manera inicial y posteriormente se documenta la ocurrencia del evento, y de esta manera el investigador puede controlar la calidad de la medición de la exposición y del evento. Los estudios considerados mixtos o ambispectivos son una mezcla en la que parte de los eventos se registra de manera prospectiva y otra parte de manera retrospectiva.

En relación con los criterios de elegibilidad que se utilizan para determinar la población de estudio, las cohortes pueden ser fijas (o cerradas) o dinámicas (o abiertas) (figura 1). Las cohortes cerradas o fijas son aque-



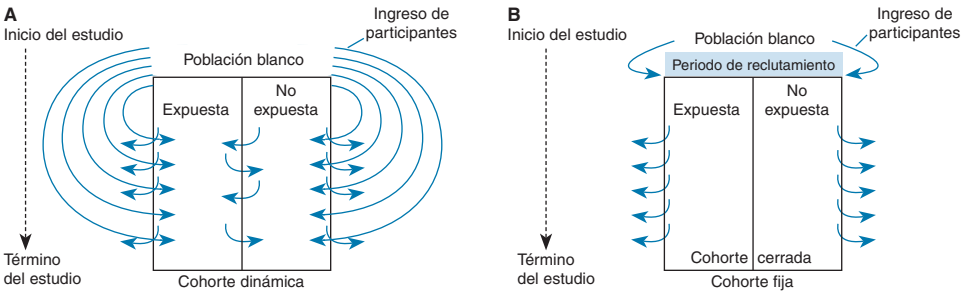



**Figura 1** **Clasificación de estudios de cohorte**

llas en las que el diseño del estudio no considera la inclusión de sujetos de observación más allá del periodo de reclutamiento fijado por los investigadores; en este tipo de estudios la población tiende a decrecer a lo largo del tiempo. Un ejemplo clásico es la cohorte de médicos británicos que fue observada a lo largo de 50 años, y donde particularmente se evalúo la mortalidad por cáncer atribuida a la exposición a tabaquismo y que finalizó con el registro de muerte del último participante del estudio de referencia.[1] Otro estudio digno de destacar es la cohorte de mujeres (Nurse's Health Study I) que en EUA fue reclutada en 1976, cuando 121 700 enfermeras entre 30 y 55 años retornaron un cuestionario inicial.[2] Este grupo de investigación inició una cohorte de mujeres jóvenes (NHS II) en 1989, con 116 671 enfermeras entre 25 y 42 años;[3] así como una nueva cohorte de 51 529 hombres, todos ellos profesionales de la salud entre 40 y 75 años, reclutados en el año de 1986.[4]

Las cohortes dinámicas o abiertas son aquellas que permiten la entrada y salida de nuevos sujetos de estudio o de sujetos que previamente habían salido de la cohorte, lo que puede ocurrir en cualquier momento, mientras esté en curso el estudio. En este tipo de cohortes el número de sujetos en estudio, en un punto cualquiera en el tiempo, puede variar o permanecer constante. Los participantes entran, reingresan o salen de la cohorte cuando cumplen con los criterios de elegibilidad establecidos por los investigadores, y contribuyen al denominador no como individuos sino con el tiempo-persona que permanecen en la cohorte en riesgo de padecer el evento y en caso de ocurrir este último, contribuyen al numerador como eventos. Este tipo cohortes es difícil de llevar a cabo por los costos que implica el tener un seguimiento individualizado de todos los miembros de la cohorte, por lo que en las mejores circunstancias deben de ser realizados en unidades geográficas y/o poblacionales en las que existe un sistema de registro de pertenencia a la cohorte. Un ejemplo de esta modalidad de estudio de cohorte fue la evaluación de los factores predictores de mayor sobrevida de los ancianos griegos en tres poblaciones, para lo cual se conformó una cohorte prospectiva abierta de 182 sujetos del sexo masculino con inicio del reclutamiento entre 1988 y 1990, así como el final del periodo de seguimiento entre 1993 y 1994. Los sujetos que reunieron criterios de inclusión en las tres poblaciones fueron incorporándose a la cohorte hasta el final del periodo de seguimiento. Se eligió este diseño de estudio porque se disponía de datos brutos y porque la incidencia de mortalidad es relativamente alta entre personas mayores, lo que permite una potencia estadística incluso en series pequeñas en la que los cálculos requeridos son sencillos.[5]



## DISEÑO DE UN ESTUDIO DE COHORTE

Para reclutar a un individuo potencialmente elegible de participar en un estudio de cohorte, se requiere que, al iniciar la investigación, la persona no manifieste el evento de estudio, pero se encuentre en riesgo de presentarlo. Una vez seleccionados los sujetos y que han dado su consentimiento para participar en el estudio, se clasifican de acuerdo con la exposición. Posteriormente, se observan a lo largo del tiempo (en forma retrospectiva o prospectiva) para cuantificar los que manifiestan el resultado.

Los dos grupos de comparación (expuestos y no expuestos) pueden seleccionarse de poblaciones diferentes; sin embargo, la posibilidad de establecer inferencia causal (la vali-

**Cuadro I Ventajas y desventajas de los estudios de cohorte**

*Ventajas*

- Permiten establecer directamente la incidencia
- La exposición se determina antes de que se conozca el resultado; es decir, existe una clara secuencia temporal entre la exposición y el evento de interés
- Brindan la oportunidad para estudiar exposiciones poco frecuentes
- Permiten evaluar resultados múltiples (riesgos y beneficios) que podrían estar relacionados con una exposición
- La incidencia del evento de interés puede determinarse para los grupos de expuestos y no expuestos
- No es necesario dejar de tratar a un grupo, como sucede con el ensayo clínico aleatorizado

*Desventajas*

- Pueden ser muy costosos y requerir mucho tiempo, particularmente cuando se realizan de manera prospectiva
- El seguimiento puede ser difícil y las pérdidas durante ese periodo pueden influir sobre los resultados del estudio
- Los cambios de la exposición en el tiempo y los criterios de diagnóstico pueden afectar a la clasificación de los individuos
- Las pérdidas en el seguimiento pueden introducir sesgos de selección
- Se puede introducir sesgos de información, si la identificación del evento puede estar influida por el conocimiento del estado de exposición del sujeto
- No son útiles para enfermedades poco frecuentes, porque se necesitaría un gran número de sujetos
- La disponibilidad de resultados depende del tiempo de ocurrencia del evento de interés
- Evalúan la relación entre evento del estudio y la exposición a sólo un número relativamente pequeño de factores cuantificados al inicio del estudio

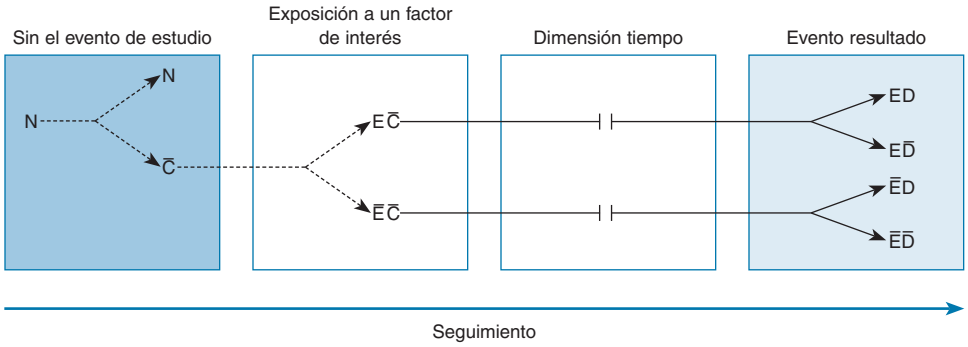


dez del estudio) depende del supuesto de que ambos grupos sean equivalentes respecto a la presencia de otros factores asociados con la exposición y el evento de interés. La hipótesis de equivalencia se cumple si en el evento téorico de poder invertir la condición de exposición entre los grupos, es posible suponer que se observaría el mismo resultado. Una de las fortalezas de los estudios de cohorte es que tienen una relación causa efecto en la dirección apropiada, ya que el comportamiento del factor de estudio se observa a través del periodo de seguimiento antes de que se detecte la enfermedad. Consecuentemente, el investigador puede postular razonablemente la hipótesis de que la exposición precede a la ocurrencia del evento y que el estatus de la enfermedad no influyó diferencialmente en la selección o seguimiento de los sujetos de estudio, ni en la medición de la exposición. Los estudios de cohorte tienen ciertas ventajas y desventajas respecto a otro tipo de estudios epidemiológicos (cuadro I) pero, en general, son menos susceptibles de presentar sesgos de selección y de información, como se describe más adelante.

Como puede observarse en la figura 2, la información acerca de la exposición se conoce para todos los sujetos al inicio del periodo de seguimiento. A partir de este momento, se observa a la población para determinar ya sea los cambios en la exposición o la ocurrencia del evento durante un periodo de tiempo determinado; esto último implica la actualización de la información mediante nuevos exámenes o cuestionarios en todos los miembros de la cohorte que se encuentran en riesgo.

### Selección de la cohorte

Antes de que puedan identificarse los posibles participantes en relación a la condición de exposición, es necesario definir explícitamente los niveles y duración mínima de la exposición; adicionalmente, en el estudio pueden fijarse criterios de elegibilidad. No obstante, lo




N = población de estudio
C = casos prevalentes
D = casos incidentes
C̄ o D̄ = no casos
E = población expuesta
Ē = población no expuesta
⊣ ⊢ = por lo menos dos mediciones en el tiempo

**Figura 2** **Diseño clásico de un estudio de cohorte**

más importante es que los sujetos (ya sea que se clasifiquen como expuestos o como no expuestos) estén en riesgo de padecer el evento de interés al iniciar el estudio.

En la medida de lo posible, los sujetos no expuestos o grupo de referencia deben ser similares a los expuestos en todos los aspectos, excepto en que no presentan la exposición que se estudia; deben tener, además, el mismo riesgo potencial de presentar el evento, y tener las mismas oportunidades que los expuestos para ser diagnosticados/detectados en relación con el evento resultado en estudio.

Las opciones utilizadas para la conformación de los grupos de la cohorte varían según la exposición que se investiga. Así, por ejemplo, el estudio de la dieta o estilos de vida se ha realizado clásicamente en cohortes formadas con muestras de la población general, mientras que exposiciones poco frecuentes, como el estudio de la exposición a asbesto, se han estudiado en cohortes ocupacionales.

La exposición y las covariables en estudio pueden ser fijas o pueden cambiar en el tiempo. En este último caso es necesario tomar en cuenta los cambios en la exposición, así como la posibilidad de que los confusores y modificadores de efecto también varíen. En las exposiciones fijas, los factores no cambian, como sería el caso del sexo, el peso al nacer, o la composición genética, entre otros.

Debido a la participación que tiene la dimensión tiempo en los estudios de cohorte, es necesario considerar que la exposición puede experimentar cambios a lo largo del seguimiento, y que la duración misma de la exposición puede reflejar un índice de exposición acumulada de relevancia biológica, como los años de duración de exposición al cigarrillo. De igual manera, el tiempo transcurrido desde la exposición útil al evento refleja un periodo de latencia; finalmente, el tiempo desde el final de la exposición puede reflejar declinación en el riesgo y también ser de interés biológi-

co, como también ocurre con la exposición a tabaco. Existen otras expresiones de la exposición que pueden estudiarse en los estudios de cohorte cuando se cuenta con mediciones y actualizaciones de la condición de exposición (cuadro II).[6]

**Medición del evento de interés**

Los eventos de estudio pueden ser:

a) evento fijo (muerte, enfermedad no recurrente —diabetes mellitus, artritis reumatoide—, eventos fisiológicos —seropositividad para VIH, cambios hormonales asociados a la menopausia—): en este caso, al observar el evento en cada unidad de análisis, el seguimiento termina para los sujetos que presentan el evento, ya que no están en riesgo de desarrollarlo;
b) eventos múltiples (enfermedades recurrentes —paludismo—, sintomatología —diarrea, crisis asmática— o cambios hormonales): al presentarse el evento, el individuo deja de estar en riesgo, por lo que ya no cumple con el criterio de permanencia en la cohorte, aunque puede reiniciarse el seguimiento cuando se restablece el riesgo, es decir, cuando hay curación y el individuo vuelve a estar en riesgo de presentarlo;
c) cambio en una medida de función (función broncopleural en el tiempo, modificación de la función pulmonar hacia un aumento o disminución, etc.), la cual se evalúa mediante tasa de cambio, y
d) marcadores intermedios del evento (i.e. concentración de apolipoproteínas A y B como marcadores de predisposición a enfermedad cardiovascular).

El periodo de seguimiento puede abarcar años, meses, semanas o días, dependiendo de la incidencia y características del evento estudiado. Al menos dos momentos definen el periodo de seguimiento: el examen inicial (medición basal) y el examen final del seguimiento. El inicio del seguimiento depende de si la cohorte es cerrada o dinámica, ya que, en el caso



**Cuadro II Aplicaciones de los estudios de cohorte**

| *Exposición de interés* | *Aplicación* | *Ejemplo* |
|---|---|---|
| Edad | Tabla de vida | Esperanza de vida de los individuos de 70 años |
| Fecha de nacimiento | Efecto de cohorte | Tasa de cáncer cervical para mujeres nacidas en 1910 |
| Exposición | Evaluación de un factor de riesgo | Tabaquismo y cáncer de pulmón |
| Intervención preventiva | Evaluación de una medida preventiva | Disminución de la incidencia de cáncer de hígado después de la vacunación contra la hepatitis B |
| Enfermedad | Evaluación de un factor pronóstico | Tasa de supervivencia de mujeres con cáncer cervical |
| Intervención terapéutica | Evaluación de un tratamiento | Supervivencia de mujeres con cáncer de ovario epitelial unilateral, a quienes se practicó *cirugía conservadora* |

Modificado de Nieto GJ[6]

de esta última, el inicio del seguimiento se define para cada participante en la medida que entra y sale de la cohorte, es decir, en función de los criterios de elegibilidad. Para el caso de las cohortes fijas con exposiciones que cambian en el tiempo también es frecuente que se lleven a cabo actualizaciones tanto de la ocurrencia del evento como de los cambios potenciales en la condición de exposición.

El seguimiento, dependiendo del evento de interés, puede ser activo o pasivo. En el primero se utilizan contactos repetidos por diversos medios: nueva entrevista y obtención de muestras, cuestionarios autoaplicables o llamadas telefónicas. El segundo se realiza mediante la búsqueda sistemática de información en registros preestablecidos (registros de cáncer, hospitalarios, registro civil, entre otros).

**Pérdidas en el seguimiento**

Las pérdidas en el seguimiento pueden originarse principalmente por tres razones: a) abandono del estudio o declinación de los sujetos de estudio para continuar, b) muerte (o muerte por otra causa, si ésta era el evento de interés), y c) pérdidas llamadas "administrativas", originadas por la terminación temprana del estudio, debido a razones ajenas a las planteadas originalmente (ejemplo: agotamiento de la fuente de financiamiento). Es importante estudiar las causas que producen pérdidas en el seguimiento para evaluar el posible impacto de éstas en la validez del estudio. Posteriormente discutiremos a detalle el impacto de las pérdidas de seguimiento en función de la validez de los estudios de cohorte.

## ANÁLISIS ESTADÍSTICO DE UN ESTUDIO DE COHORTE



La base del análisis de un estudio de cohorte es la evaluación de la posible incidencia diferencial (en exceso o déficit) de un evento durante el seguimiento en el tiempo de los sujetos en riesgo, como consecuencia de haber estado expuesto o no a una determinada exposición. Esto es, el investigador selecciona un grupo de sujetos expuestos y otro de no expuestos y los sigue en el tiempo para comparar la incidencia de algún evento (incidencia de la enfermedad o, según sea el caso, tasa de muerte de la enfermedad o, incluso, curación). Es indispensable considerar que para poder analizar adecuadamente un estudio de cohorte, es necesario tener la fecha de inicio del seguimiento, la fecha en que ocurren los eventos y la de término del seguimiento, así como la información completa sobre los sujetos participantes, la escala de medición y el motivo de terminación del seguimiento (pérdida, muerte u ocurrencia del evento en estudio).

**Razón de incidencia acumulada**

Cuando existe una asociación positiva entre la exposición y el evento, se esperaría que la proporción del grupo expuesto que lo desarrolló sea mayor que la proporción del grupo no expuesto (incidencia del grupo expuesto *vs.* incidencia del grupo no expuesto). Partiendo de un grupo expuesto, donde "a" sujetos desarrollan el evento y "c" sujetos no desarrollan el evento, tenemos entonces que la incidencia de la enfermedad entre los expuestos es: a / a + c. De la misma manera, en el grupo de sujetos no expuestos ("b" y "d") el evento ocurre en "b" sujetos, pero no en "d" sujetos; tenemos entonces que la incidencia de la enfermedad entre los no expuestos es: b / b + d (cuadro III).

**Cuadro III Análisis de un estudio de cohorte para evaluar razón de riesgos**

| | | Exposición | |
|---|---|---|---|
| | | Sí | No |
| Evento | Sí | a | b |
| | No | c | d |
| | Total | $m_1$ | $m_0$ |

$IA_1$ = incidencia acumulada en expuestos = a/(a + c)

$IA_0$ = incidencia acumulada a no expuestos = a/(b + d)

Razón de incidencia acumulada = $IA_1/IA_0$

Diferencia de incidencia acumulada* = $IA_1 - IA_0$

*Si la exposición es protectora, la diferencia de riesgos debe calcularse como $IA_0 - IA_1$

donde: a = Sujetos con la exposición que desarrollan el evento
b = Sujetos sin la exposición que desarrollan el evento
c = Sujetos con la exposición que no desarrollan el evento
d = Sujetos sin la exposición que no desarrollan el evento
$m_1$ = Total de sujetos expuestos
$m_0$ = Total de sujetos no expuestos



Para calcular la razón de incidencia acumulada (RIA), se divide la incidencia del grupo expuesto entre la del grupo no expuesto:

$$RIA = \frac{IA_1}{IA_0}$$

La RIA, también conocida como razón de riesgos o riesgo relativo, es una medida de asociación entre el evento y la exposición. Consideremos un ejemplo con datos hipotéticos de un estudio de cohorte en el que se investiga la asociación entre el estado nutricional y la muerte en pacientes con diagnóstico de leucemia. Se seleccionó un grupo de 17 sujetos con estado nutricional bajo (expuestos) y otro de 15 sujetos con estado nutricional normal (no expuestos), quienes se encontraban libres de la enfermedad al inicio del estudio. A ambos grupos se les dio seguimiento hasta que se presentó el evento. El evento (muerte) se registró en 14 sujetos del grupo con déficit nutricional y en ocho en el grupo sin déficit. El resultado es una incidencia de 0.82 en el grupo de bajo estado nutricional y de 0.53 en los del grupo de estado nutricional normal. Al estimar el efecto obtenemos un riesgo relativo mayor de uno; por lo tanto, los sujetos con bajo estado nutricional tienen 1.54 veces el riesgo de presentar el evento, comparados con los sujetos con estado nutricional normal (cuadro IV).

### Razón de tasas de incidencia

En un estudio de cohorte se presentan tiempos de seguimiento desiguales, ya sea por los diferentes momentos de aparición del evento de interés o por pérdidas en el seguimiento (abandono, cambio de domicilio, muerte, fina-

**Cuadro IV Cálculo de la razón de riesgos para sujetos con bajo estado nutricional en relación con leucemia**

| | | Estado nutricional | | |
|---|---|---|---|---|
| | | Bajo | Normal | Total |
| Leucemia | Sí | 14 | 8 | 22 |
| | No | 3 | 7 | 10 |
| | Total | 17 | 15 | 32 |

Incidencia en el grupo de expuestos ($m_1$) = a/(a + c) = 0.82

Incidencia en el grupo de no expuestos ($m_0$) = b/(b + d) = 0.53
Razón de incidencia acumulada = 0.82/0.53 = 1.54

Diferencia de incidencia acumulada = 0.82 – 0.53 = 0.29

IC 95% para la razón de incidencia acumulada* = 0.91 – 2.6

* IC 95% = $e^{\ln RIA \pm 1.96 * \sqrt{(c/am + d/bm)}}$

lización del estudio por agotamiento de recursos financieros). Una forma de tratar periodos de seguimiento variables es mediante el análisis basado en tiempo-persona. En estos casos, puede utilizarse el promedio de tiempo contribuido por la totalidad de sujetos de la cohorte.[7] Cuando se ha cuantificado el tiempo-persona de seguimiento para cada sujeto, el denominador cambia a una dimensión de tiempo (las unidades son, por ejemplo, años-persona, días-persona, horas-persona). Esto nos permite estimar la tasa de los casos incidentes en una unidad de tiempo determinada.[8]

Consideremos una cohorte en la que se conoce el tiempo que cada individuo ha permanecido en el seguimiento. Puede calcularse la tasa de incidencia de acuerdo con el estado de exposición de cada uno. Así, partiendo de un grupo de sujetos "a" que presentan el evento y la exposición, y el tiempo-persona de seguimiento "$tp_e$" de estos sujetos expuestos, puede calcularse la tasa de incidencia para los expuestos: $TI_1 = a/tp_e$. De la misma manera, consideremos un grupo "b" de sujetos con el evento, pero sin la exposición y el tiempo-persona de seguimiento de estos sujetos "$tp_{ne}$", en los cuales se puede calcular la tasa de incidencia para los sujetos no expuestos: $TI_0 = b/tp_{ne}$ (cuadro V). El cálculo de la razón de tasas de incidencia (RTI) se realiza de la siguiente manera:

$$RTI = TI_1/TI_0$$

El resultado de esta estimación es una medida de asociación que nos permite comparar los riesgos entre los grupos expuesto y no expuesto.

Tomando el mismo ejemplo hipotético de estado nutricional y leucemia, en el cual 14 sujetos con bajo estado nutricional (contribución de 571 días-persona de seguimiento) y ocho sujetos del grupo con estado nutricional normal (contribución de 1 772 días-persona de seguimiento) desarrollaron el evento (muerte), se observa que la tasa de mortalidad para

**Cuadro V Análisis de un estudio de cohorte para estimar razón de tasas de incidencia**

| | Expuestos | No expuestos |
|---|---|---|
| Casos | a | b |
| Tiempo-persona | $tp_e$ | $tp_{ne}$ |
| Tasas | $TI_1$ | $TI_0$ |

Tasa de incidencia en el grupo de expuestos ($TI_1$) = $a/tp_e$

Tasa de incidencia en el grupo de no expuestos ($TI_0$) = $b/tp_{ne}$

Razón de tasas de incidencia = $TI_1/TI_0$

Diferencia de tasas de incidencia* = $TI_1 - TI_0$

*Si la exposición es protectora, la diferencia de tasas debe calcularse como $TI_0 - TI_1$

donde:
- a = Sujetos con la exposición que desarrollaron el evento
- b = Sujetos sin la exposición que desarrollaron el evento
- $tp_e$ = Tiempo-persona de seguimiento de los sujetos expuestos que desarrollaron el evento
- $tp_{ne}$ = Tiempo-persona de seguimiento de los sujetos no expuestos que desarrollaron el evento
- $TI_1$ = Tasa de incidencia de los sujetos expuestos
- $TI_0$ = Tasa de incidencia de los sujetos no expuestos



los sujetos con bajo estado nutricional fue de 24.5/1 000 días-persona y para los sujetos con estado nutricional normal, de 4.5/1 000 días-persona. Por lo tanto, la velocidad de ocurrencia del evento es 5.4 veces más alta en el grupo de sujetos con estado nutricional bajo que en el grupo de sujetos con estado nutricional normal (cuadro VI). Como puede verse, los resultados de la estimación de la razón de riesgos y la razón de tasas de incidencia es diferente; esto se debe a que el evento es frecuente (22 de 32 sujetos lo presentaron) y el periodo de seguimiento es prolongado. En esta condición es esperable que ambas medidas sean discrepantes, ya que la suposición de que todos los individuos contribuyeron de igual manera en el caso de la estimación de la RIA es incorrecta.

Este método permite realizar análisis cuando existe un cambio en el estado de exposición, de tal manera que un mismo sujeto puede contribuir en el denominador de los expuestos en un periodo y entre los no expuestos en otro.

Finalmente, aunque hemos descrito de manera simple la forma como se analizan los datos provenientes de una cohorte, cuando queremos conocer el efecto de la variable estudiada controlando variables potencialmente confusoras, requerimos de un análisis multivariado ajustando simultáneamente diferentes variables mediante técnicas estadísticas específicas, como la regresión de Poisson[7] o el análisis de supervivencia.[9] Esta última estrategia permite el análisis de eventos frecuentes en poblaciones pequeñas, lo que permite también una estimación mas apropiada de la RIA, a dife-

**Cuadro VI** **Cálculo de tasas de incidencia de muerte por leucemia de acuerdo a estado nutricional**

| | Estado nutricional | |
|---|---|---|
| | Bajo | Normal |
| Leucemia | 14 | 8 |
| Tiempo persona* | 571 | 1772 |
| Tasas | 0.0245 | 0.0045 |

Tasa de incidencia en el grupo de expuestos ($TI_1$) = $a/tp_e$ = 0.0245

Tasa de incidencia en el grupo de no expuestos ($TI_0$) = $b/tp_{ne}$ = 0.0045
Razón de tasas de incidencia = $TI_1/TI_0$ = 5.4

Diferencia de tasas de incidencia acumulada = $TI_1 - TI_0$ = 0.02

IC 95% para la razón de tasas de incidencia acumulada† = 0.91 - 2.6

* días-persona
† IC 95% = $e^{\ln RTI \pm 1.96 * \sqrt{(1/a + 1/b)}}$

rencia del análisis de tiempo-persona, en el que el evento es generalmente poco frecuente y se realiza en estudios que involucran tamaños muestrales mayores. El cuadro VII presenta las principales diferencias entre las dos estrategias de análisis.[6]



**Cuadro VII** **Comparación de las principales estrategias para el análisis de los estudios de cohorte***

| *Tamaño de muestra* | *Análisis de supervivencia*<br>*Relativamente pequeño* | *Análisis basado en tiempo-persona*<br>*Relativamente grande* |
|---|---|---|
| Tipo de eventos | Frecuentes | Raros |
| Escala temporal | Única | Única o múltiple |
| Tipos de medida | Probabilidad (condicional y acumulada) | Tasa (densidad) |
| Análisis unifactorial | • Comparación de curvas de supervivencia<br>• Prueba de log rank<br>• Razón de riesgos | • Comparación de tasas<br>• Razón de tasas (densidades)<br>• Razón estandarizada de mortalidad (REM) |
| Análisis multifactorial | Regresión de Cox | Regresión de Poisson |

* Modificado de Nieto GJ[4]

## SESGO Y VALIDEZ EN LOS ESTUDIOS DE COHORTE

Aunque se reconoce que los estudios de cohorte representan un diseño menos sujeto a error sistemático o sesgo en comparación con otros estudios observacionales,[10] no es menos cierto que deben tenerse en consideración la posibilidad de presentarse posibles sesgos que pueden distorsionar los resultados de ellos derivados (figuras 3-5). Existen, en efecto, sesgos de selección e información en los estudios de cohorte que deben considerarse rigurosamente, sobre todo en lo que se refiere a pérdidas en el seguimiento (de los pacientes, participantes, etc.), al modo en que se obtiene la información sobre la exposición estudiada y al modo en que se determina en la población en estudio la ocurrencia de la enfermedad o condición de interés durante el seguimiento. Por lo que se refiere a sesgos de confusión, en los estudios de cohorte es importante considerar factores que se asocien independientemente, tanto con la exposición como con la condición o evento estudiado, y que no sean pasos intermedios en la cadena causal, ya que la presencia diferencial de estos factores entre los grupos expuesto y no expuesto, pueden hacer aparecer una asociación ficticia entre la exposición y el evento en



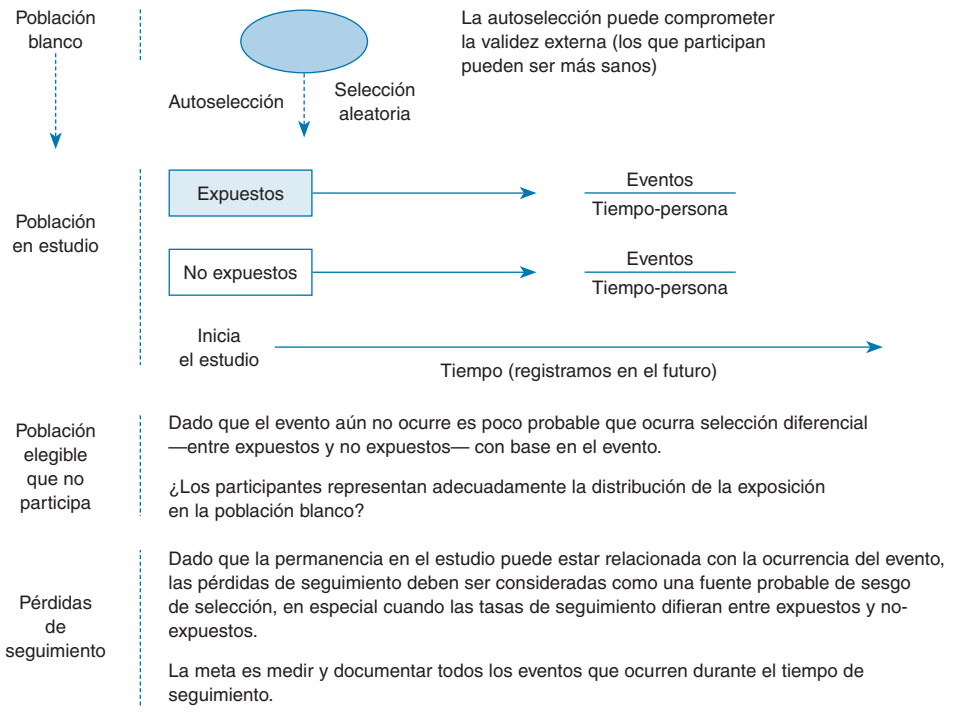



**Figura 3** **Sesgos de selección en estudios de cohorte prospectivos. Población elegible que no participa**

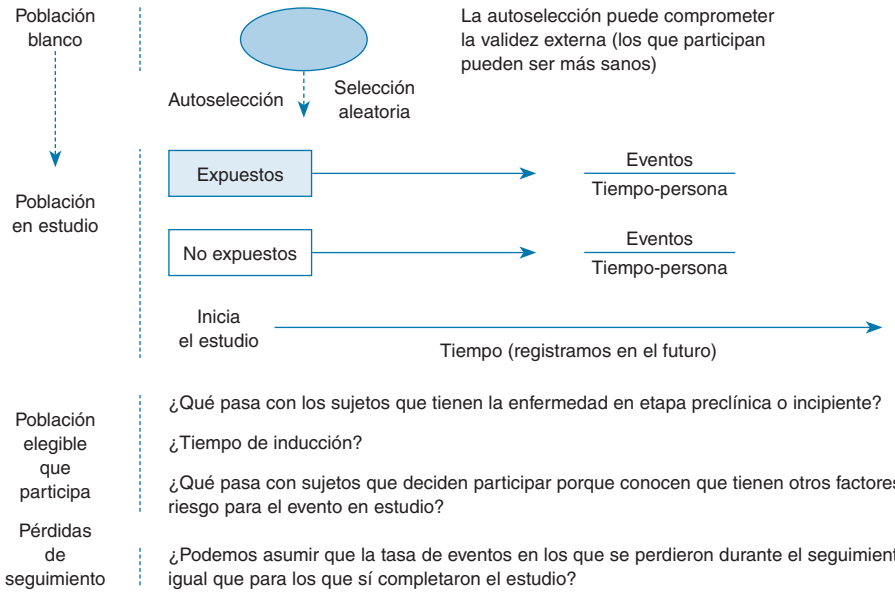



**Figura 4** **Sesgos de selección en estudios de cohorte prospectivos. Población elegible que participa**



estudio.[11] El procedimiento para la identificación y el control de la confusión (mediante el pareamiento, análisis estratificado y modelaje multivariado, principalmente) es conceptualmente similar al usado en otros diseños epidemiológicos. Centraremos el resto de la discusión en las principales fuentes de sesgos.

Los sesgos en los estudios epidemiológicos se han clasificado tradicionalmente como sesgos de selección (cuando los errores se derivan de la constitución de la población en estudio) y sesgos de información (cuando los errores se originan durante el proceso de recolección de la información).

### Sesgos de selección

Los sesgos de selección en una cohorte tienen que ver tanto con la validez interna como con la validez externa o extrapolación de los resultados que se obtengan. Este tipo de sesgos está relacionado, evidentemente, con el procedimiento utilizado para conformar la cohorte o población en estudio: cuando la población en estudio se constituye con voluntarios, la representatividad que este grupo pueda tener de la población blanco (de la población a la cual se pretende generalizar los resultados) puede estar limitada por el hecho de que los voluntarios sean diferentes en algunos aspectos a la población general. Por ejemplo, en el *Estudio europeo sobre dieta y cáncer,*[12] la cohorte española se conformó a partir de donadores de sangre, es decir, hombres y mujeres altruistas, mayoritariamente jóvenes, a pesar de lo cual la validez interna de los resultados no se vería afectada. ¿Cómo puede constituirse una cohorte representativa de una determinada población? Un diseño experimentado con cierto éxito es el de

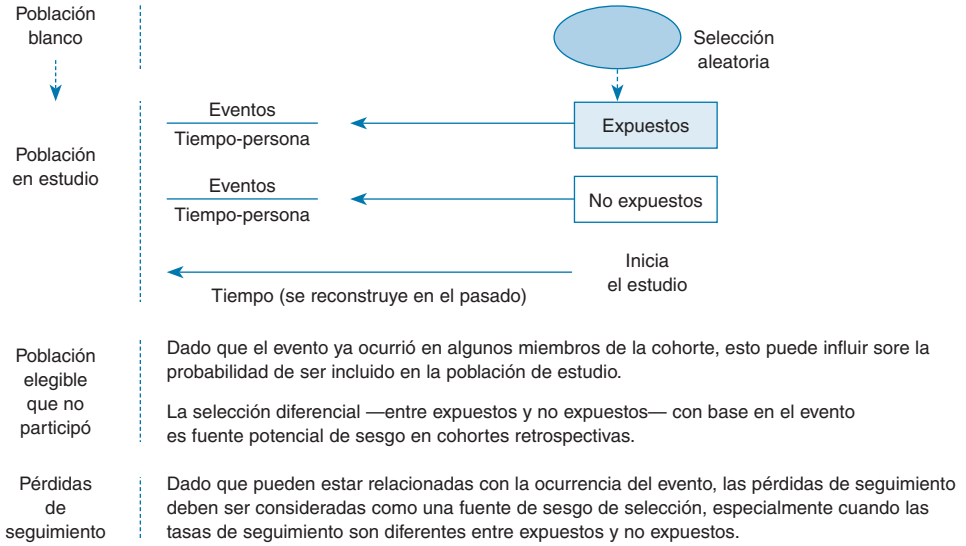



**Figura 5 Sesgos de selección en estudios de cohorte retrospectivos. Población elegible que no participó**



seguimiento de muestras representativas de la población general. Un ejemplo de este tipo de diseño es el estudio de Sepúlveda y colaboradores,[13,14] basado en el seguimiento activo de cerca de 3 241 niños de la Delegación Tlalpan, en la Ciudad de México, que permitió hacer la investigación prospectiva del papel del estado nutricional como un factor de riesgo para enfermedades diarreicas, en una cohorte representativa de la población infantil de un área geográfica de esa ciudad.

A pesar de contar con una cohorte inicial representativa, ¿cómo puede asegurarse que determinadas personas que retiren su participación del estudio, ya sea en la fase inicial de recolección de información o incluso después de cierto tiempo de seguimiento, no condicionen diferencias en los grupos estudiados que conduzcan a errores en los resultados? Las pérdidas en el seguimiento son, justamente, la principal fuente de sesgo de selección, ya sea que se trate de cohortes constituidas a partir de muestras representativas de la población, voluntarios o colectivos definidos (como pueden ser los trabajadores de una determinada fábrica o sector industrial o grupos profesionales determinados como médicos, enfermeras o educadores). Las pérdidas en el seguimiento no invalidan *per se* el estudio, pero los investigadores deben utilizar procedimientos para minimizar su ocurrencia y, en caso de que se presenten, considerar si afectan o no los resultados. Para ello, es necesario tratar de recoger información clave sobre los participantes que abandonen el estudio, en especial para investigar si el abandono tiene alguna relación con las exposiciones o las enfermedades o eventos estudiados.

Un claro ejemplo de sesgo de selección que puede alterar considerablemente los resultados obtenidos es el llamado efecto del trabajador sano, observado frecuentemente en cohortes

laborales cuando el grupo no expuesto queda constituido por la población general. Este último grupo, a diferencia del grupo de trabajadores que incluye sujetos sanos, activos y con empleo, incluye a sujetos similares, además de los desempleados y a los enfermos que no pueden trabajar. Por ejemplo, al estudiar la mortalidad por cáncer y por enfermedades cerebrovasculares de los trabajadores de una determinada empresa relacionada con productos radioactivos, y compararla con la mortalidad esperada de acuerdo con las tasas observadas en la población general, se observa que el riesgo relativo de muerte por esas causas es de 0.78 y 0.81, ambos con intervalos de confianza que no incluyen la unidad.[15] Los resultados parecerían sugerir que el trabajo en esa empresa protege a sus empleados con una disminución de la frecuencia con la que se presentan las enfermedades estudiadas. ¿Cómo explicar estos resultados? La cohorte de trabajadores representa un grupo especial de personas de la población general: son personas en edad laboral y con capacidad física óptima, existe una autoselección de las personas que van a desarrollar tareas concretas en esa fábrica, de manera que personas con determinados trastornos crónicos se autoexcluyen (o son excluidos por la dirección de la empresa) o pasan a trabajos menos pesados. Así, al comparar la mortalidad en ese grupo de trabajadores sanos con el experimentado en la población general del país, la observada en la empresa debe ser inferior. Una manera de evitar este tipo de sesgo de selección es realizando comparaciones internas en el seno de la cohorte (por ejemplo, la mortalidad por cáncer en los trabajadores de una misma empresa, pero expuestos a diferentes niveles de radiación) con lo que puede eliminarse en buena parte el sesgo del trabajador sano.[16] Los estudios de cohorte retrospectiva son más vulnerables a los sesgos de selección. Esto se debe a que, al inicio de este tipo de trabajos, el evento ya ocurrió en un buen número de los participantes, y esto puede influir en la probabilidad de formar parte del estudio. Este tipo de sesgo en estudios retrospectivos es particularmente serio cuando los participantes conocen también su condición de exposición y cuando la presencia conjunta de estos eventos (exposición y enfermedad) motiva una participación diferencial en el estudio.

Otro tipo de sesgo de selección, acaso el más problemático y frecuente en los estudios de cohorte, se debe a pérdidas en el seguimiento, las cuales estarán sesgando los resultados si se encuentran relacionadas con alguna característica de los participantes, como puede ser la misma exposición o el desenlace estudiado. En este caso, como en el del efecto del trabajador sano, el sesgo de selección introducido por las pérdidas durante el seguimiento compromete la validez interna del estudio, es decir, los grupos expuesto y no expuesto no son comparables, por lo que se pierde la veracidad de los resultados. Así, las pérdidas deben ser independientes de la condición de exposición, es decir, se deben presentar con la misma frecuencia en los grupos expuesto y no expuesto y los sujetos que se pierden deben ser de características similares, misma edad y sexo, entre otros. Uno de los motivos más frecuentes de pérdidas en el seguimiento obedece a la movilidad de los participantes de la cohorte, que dificulta los contactos repetidos que se utilizan tanto para realizar nuevas determinaciones de las exposiciones como para conocer el desarrollo de las condiciones estudiadas (incidencia de enfermedades o muerte). Para mantener y maximizar la participación, y minimizar las pérdidas en el seguimiento, es necesario poner en práctica diferentes estrategias durante el desarrollo del estudio.[17]

Un ejemplo de estudio de cohorte en el que se han invertido grandes esfuerzos en el seguimiento es el de las enfermeras estadounidenses, el cual comenzó en 1976, con la inclusión de 121 700 enfermeras certificadas, en edades

comprendidas entre 30 y 55 años, que contestaron un cuestionario postal sobre estilos de vida y condiciones médicas.[18] Su objetivo primario era investigar la relación entre el consumo de anticonceptivos orales y el cáncer de mama, y después de más de 25 años de seguimiento ha generado múltiples resultados en relación con ése y otros numerosos problemas de salud. ¿Cómo han conseguido los investigadores realizar un seguimiento continuado de más de 100 000 participantes?

1. El diseño inicial se centró en enfermeras certificadas, de manera que aunque cambien de domicilio y de localidad de residencia, si siguen ejerciendo su profesión, seguirán activas y podrán ser localizadas.
2. En el cuestionario inicial se solicitó, además del nombre de la participante, su número de Seguro Social, fecha de nacimiento, dirección y el número de teléfono de un contacto personal.
3. Además, gran parte del esfuerzo se concentra en el seguimiento mediante el cuestionario postal: se envían cuestionarios de seguimiento cada dos años, que van acompañados de una carta de presentación y un boletín con información actualizada de los progresos del estudio, y se actualiza la información de los contactos personales cada cuatro años. En caso de que no haya respuesta, el cuestionario se envía hasta cinco veces (la quinta vez se trata de una versión abreviada).
4. Tras el quinto envío postal se realiza un seguimiento telefónico; se utiliza correo certificado o de mensajería privada; se consulta a los carteros locales, los colegios de enfermería y los contactos personales.
5. Finalmente, una parte importante del seguimiento de la mortalidad se realiza mediante el uso del Registro Nacional de Defunciones estadounidense,[19] que consiste en un índice sistematizado de todas las muertes acaecidas en los Estados Unidos desde 1979. De esta manera, desde 1990, mediante una combinación de estrategias, la cohorte de las enfermeras cuenta con 90% de seguimiento.



**Sesgos de información**

La introducción de errores sistemáticos que comprometan la validez interna del estudio por el modo en que se obtuvo la información o los datos de los participantes se conoce como sesgo de información. Este tipo de sesgo se presenta con frecuencia cuando la información se obtiene de manera diferente en los grupos estudiados; por ejemplo, cuando los participantes en el grupo expuesto son seguidos, monitorizados o vigilados de manera más cuidadosa que los participantes en el grupo no expuesto. En este mismo sentido, en estudios clínicos de seguimiento, es frecuente que algunos participantes presenten condiciones de comorbilidad que generen, incluso de manera no apreciable para los investigadores, una mayor vigilancia o control de esos pacientes en relación con otros sujetos del estudio, lo que aumenta artificialmente las posibilidades de diagnóstico de la condición de interés. Esto será una fuente de sesgo siempre y cuando ocurra de manera diferencial entre expuestos y no expuestos. Por ejemplo, entre las participantes posmenopáusicas, en la cohorte de las enfermeras estadounidenses es más probable que las que toman terapia sustitutiva hormonal se sometan a un seguimiento más riguroso que, acaso, incluya prácticas preventivas adicionales como la detección oportuna de cáncer de mama, lo que haría más probable la detección de tumores mamarios en este grupo de la cohorte, en comparación con las participantes posmenopáusicas que no siguen un tratamiento sustitutivo y llevan un seguimiento médico y ginecológi-

co habitual.[20] Así, al cuantificar la asociación entre la terapia hormonal sustitutiva y cáncer mamario, se observaría un incremento de la incidencia de la enfermedad, debido al sesgo de información (mayor detección de cáncer de mama) entre las participantes de la cohorte en tratamiento sustitutivo.

En ocasiones, es el propio investigador quien evalúa de forma sesgada la presencia o no de la condición de interés, puesto que conoce las hipótesis bajo investigación o la historia de exposiciones de los participantes. Este tipo de error se conoce como sesgo del observador. Imaginemos que en el estudio mencionado anteriormente "estamos convencidos" de que la terapia sustitutiva aumenta el riesgo de cáncer mamario, por lo que podríamos incurrir en un sesgo del observador si se intensificara el seguimiento de las enfermeras que toman la terapia sustitutiva, o alternativamente, podríamos minimizar el seguimiento de las enfermeras que no consumen hormonas. Lo mismo podría ocurrir si el radiólogo o patólogo que lleva a cabo el diagnóstico, lo hace considerando la información sobre el uso de terapia de reemplazo. El problema se deriva de realizar un esfuerzo o decisión diferente en el seguimiento de los expuestos y los no expuestos, derivado de las hipótesis a priori —o en algunos casos de la propia organización logística— del estudio. ¿Cómo puede evitarse el sesgo de información? Garantizando que todas las mediciones realizadas (mediante cuestionarios y muestras biológicas) se realicen con el mismo grado de error (misma sensibilidad y especificidad) en el grupo expuesto y no expuesto. En algunos estudios, esto se logra cegando a los participantes y a los observadores sobre la condición de exposición y la hipótesis de estudio.

**Sesgos por mala clasificación**

En los estudios de cohorte deben tenerse en cuenta los sesgos de información debidos a la clasificación errónea (mala clasificación) de los participantes respecto a la existencia o a la cuantificación de la exposición estudiada o a la ocurrencia de la enfermedad o de la condición de interés.[21] ¿De qué depende este tipo de error? La principal fuente de sesgo deriva de los instrumentos utilizados y su modo de aplicación (cuestionarios, técnicas analíticas, biomarcadores, etcétera).

Por ejemplo, si disponemos de un cuestionario para determinar el consumo de alcohol semanal que produce estimaciones inferiores a las reales (v. g., porque no considera de forma independiente el consumo de alcohol el fin de semana de su consumo el resto de la semana) incurriremos en una infravaloración del consumo real de alcohol. Esa subestimación ¿se produce del mismo modo en los grupos expuesto y no expuesto; entre los sujetos que desarrollaron el evento y los que no? Ésa es una cuestión clave que permitirá caracterizar la mala clasificación en dos tipos: a) diferencial, cuando el error en la clasificación depende del valor de otras variables, y b) no diferencial, cuando el error no depende de las otras variables. Volvamos al ejemplo anterior. Supongamos que el cuestionario sobre consumo de alcohol es administrado por entrevistadores en el domicilio de los participantes. Si el protocolo de administración no se aplica estrictamente o ha fallado el adiestramiento de los entrevistadores, podría suceder que algunos enfaticen e investiguen hasta la última gota de alcohol consumida por el participante, sobre todo cuando se entrevista a varones de mediana edad en zonas o barrios deprimidos de la ciudad en los que el consumo de alcohol parece ser una conducta habitual. Sin embargo, al entrevistar a mujeres, también de mediana edad, de zonas más favorecidas, el entrevistador no enfatiza la metodología de la entrevista, asumiendo que los participantes de esas características no tienen un consumo de alcohol importante. Nos encontramos, pues, ante un

típico ejemplo de mala clasificación de la exposición, en este caso diferencial en lo que se refiere a las variables de nivel socioeconómico y sexo. Cuando el error afecta por igual a todos los participantes, independientemente de su exposición verdadera (los entrevistadores aplican de manera similar el cuestionario, pero recoge consumos inferiores de alcohol sistemáticamente) nos hallamos ante una mala clasificación no diferencial o aleatoria. En este último caso, el sesgo introducido tiende a subestimar la asociación real en caso de que exista, es decir cambia las estimaciones del riesgo relativo hacia la hipótesis nula. Sin embargo, la dirección del sesgo en el caso de la mala clasificación diferencial depende del tipo de exposición investigada y es, en muchos casos, impredecible.

Las posibles soluciones para minimizar los errores de medición aleatorios consisten en la validación de los instrumentos de medida utilizados, ya sean cuestionarios estructurados, pruebas psicométricas, instrumentos médicos (esfigmomanómetros, balanzas, y otros), técnicas de laboratorio, etcétera, junto con la implantación de protocolos de aplicación estrictos, previo entrenamiento y estandarización de los observadores, sobre todo cuando son muchos y de diferentes centros, así como la realización de medidas repetidas en los mismos sujetos. Finalmente, si se realiza una prueba piloto en condiciones reales y se ponen en práctica controles de calidad continuados de la información recolectada, pueden predecirse los sesgos,[22] que difícilmente se controlan en las fases de análisis del estudio.

De lo anteriormente comentado, se deduce que los sesgos pueden minimizarse con un buen diseño; que en los estudios de cohorte se ha de incluir la planificación detallada de la constitución de la cohorte y los mecanismos de seguimiento, además de los instrumentos de captura de la información. La utilización de métodos estadísticos como el análisis estratificado y los métodos de regresión múltiple nos permite controlar en la investigación algunos de los sesgos que no pudieron prevenirse durante el diseño. Un estudio libre de sesgos nos garantizará su validez interna, así como su validez externa o extrapolación.



## CONCLUSIONES

La utilización de los estudios de cohorte ha aumentado considerablemente durante los últimos años. Como parte de la revolución informática mundial, se han desarrollado avances considerables en los métodos que permiten el seguimiento eficiente y costo-efectivo de grandes y diversos grupos poblacionales, así como la aplicación en epidemiología de métodos estadísticos sofisticados que permiten prevenir, corregir y controlar diferentes sesgos y examinar las relaciones epidemiológicas en un contexto más controlado. Estos avances han permitido el desarrollo de conocimiento derivado de estudios epidemiológicos de cohorte que han tenido un gran impacto en la práctica médica; un ejemplo destacado de esto lo constituyen los datos sobre el efecto de factores ambientales[23] sobre la salud humana. Sin duda, los estudios de cohorte han modificado la percepción de los estudios observacionales y ahora se consideran una herramienta importante para el avance del conocimiento médico, en especial para el estudio de condiciones específicas que no pueden ser abordadas utilizando estudios experimentales. La continua mejoría de los registros médicos y epidemiológicos en México, la implementación de bancos de biomarcadores —como los que están se están implementando en las Encuestas Nacionales de Salud—, así como la ca-

pacidad técnica e infraestructura institucional desarrolladas en los últimos años, sin duda contribuirán a que este tipo de diseño se utilice con mayor frecuencia para el estudio de diversas exposiciones ambientales, infecciosas y nutricionales, así como las relacionadas con diversos estilos de vida en diferentes contextos geopolíticos y grupos de riesgo. Esto permitirá estudiar con mayor rigor metodológico el proceso salud-enfermedad y el impacto de los diferentes programas de intervención.

## REFERENCIAS


1. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years' observations on British doctors. Br J Cancer 2005;92:426-429.
2. Colditz GA. The Nurse' Health Study: a cohort of US women followed since 1976. J Am Med Womens Assoc 1995;50:40-44.
3. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, Rosner B, Kriska A, Willett WC. Reproducibility and validity of a self administered physical activity questionnaire. Int J Epidemiol 1994;23:991-999.
4. Colditz GA, Rimm EB, Giovannucci E, Stampfer MJ, Rosner B, Willett WC. A prospective study of parental history of myocardial infarction and coronary artery disease in men. Am J Cardiol 1991;67:933-938.
5. Trichopoulo A, Ocurrís-Blazos A, Vassilakou T, *et al.* Diet and survival of elderly Greeks: A link to the past. Am J Clin Nutr 1995;61(suppl):1346S-1350S.
6. Nieto GJ. Los estudios de cohorte. En: Martínez N, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V. Salud Pública. Madrid: McGraw-Hill / Interamericana, 1999.
7. Breslow NE, Day NE. Statistical methods in cancer research. The design and analysis of cohort studies. Vol. I. Lyon: IARC Sci Pub, 1987.
8. Rothman KJ, Greenland S. Modern epidemiology. 2a. ed. Boston: Lippincot-Raven, 1998:5-100.
9. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457-481.
10. MacMahon B, Trichopoulos D. Epidemiology. Principles and methods. 2a. ed. Boston: Little Brown and Company, 1996:165-225.
11. Rothman KJ, Greenland S. Modern epidemiology. 2a. ed. Boston: Lippincot-Raven, 1998:120-186.
12. European prospective investigation into cancer and nutrition. EPIC Group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. I. Foods. Int J Epidemiol 1997;26(Suppl 1):S91-S99.
13. Sepúlveda J, Willett W, Muñoz A. Malnutrition and diarrhea. A longitudinal study among urban Mexican children. Am J Epidemiol 1988;127:365-376.
14. Sepúlveda J. Malnutrition and infectious diseases. A longitudinal study of interaction and risk factor. Cuernavaca: Instituto Nacional de Salud Pública(Perspectivas en Salud Pública), 1990.
15. Checkoway H, Pearce N, Crowford-Brown DJ. Research methods in occupational epidemiology. Nueva York: Oxford University Press, 1989:77-80.
16. Checkoway H, Eisen EA. Developments in occupational cohort studies. Epidemiol Rev 1998; 20(1):100-111.
17. Hunt JR, White E. Retaining and tracking cohort study members. Epidemiol Rev 1998;20(1):57-70.
18. Colditz GA. The nurses' health study: A cohort of US women followed since 1976. J Am Med Women Assoc 1995;50(2):40-44.
19. Howe GR. Use of computerized record linkage in cohort studies. Epidemiol Rev 1998;20(1):112-121.
20. Franceschi S, La Vecchia C. Colorectal cancer and hormone replacement therapy: An unexpected finding. Eur J Cancer Prev 1998;7(6): 427-438.
21. Mertens TE. Estimating the effects of misclassification. Lancet 1993;342: 418-421.

22. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. Epidemiol Rev 1998;20(1):71-80.
23. Eskenazi B, Gladstone E, Berkowitz G, Drew C, Faustman E, Holland N, et al. Methodologic and logistic issues in conducting longitudinal birth cohort studies: Lessons learned from the centres for children's environmental health and disease prevention research. Environmental Health Perspectives 2005;113(10):1419-1429.

# Estudios de cohorte — Ejercicios

Las preguntas de este ejercicio se basan en el siguiente artículo.

## REFERENCIA

Sepúlveda J, Willet W, Muñoz A. Malnutrition and diarrhea. A longitudinal study among urban mexican children. Am J Epidemiol 1988;127:365-376.

## RESUMEN


### Introducción

La diarrea y la desnutrición son hallazgos frecuentes en niños de países en desarrollo. Las infecciones, particularmente la diarrea, se conocen por afectar el estado nutricional de los niños, mediante varios mecanismos. Sin embargo, es menos claro si la desnutrición predispone a contraer la diarrea. Varios de los estudios realizados con anterioridad al trabajo que aquí se presenta no mostraron ser concluyentes en cuanto a si la desnutrición correspondía a un factor de riesgo de la diarrea.

### Material y métodos

*a) población de estudio*

El estudio se realizó en la Delegación de Tlalpan, una zona periurbana situada al sur de la Ciudad de México. Se identificaron 300 niños menores de dos años de edad, distribuidos de manera balanceada en tres categorías de nutrición: normal, desnutrición leve (i.e. 75%-89% de peso para la edad de acuerdo con las tablas del Centro Nacional de Estadísticas de Salud) y desnutrición moderada (i.e. 60%-74% del peso para la edad de acuerdo con las mismas tablas). Para lograrlo, se realizó una encuesta demográfica puerta a puerta. A los niños se les pesó y midió y se asentó su edad exacta.

Cada niño menor de dos años con desnutrición moderada fue pareado por edad, sexo y lugar de residencia con un niño de cada una de las otras dos categorías, seleccionándolos de manera aleatoria de todos los potenciales pares. Se eligió, cuando mucho, un niño por casa. La muestra final fue de 284 niños, ya que 16 se mudaron de casa justo antes de iniciar el estudio.

*b) medición del estado nutricional*

A los niños se les tomó el peso y talla y se les midió la circunferencia del brazo al inicio del estudio y cada tres meses durante 52 semanas. El 86% de los casos contó con 5 mediciones. Del total, los pocos niños cuyo estado nutricional cambió de desnutrición moderada a severa recibieron tratamiento médico. Para fines estadísticos, se les clasificó dentro del grupo de desnutrición moderada.

*c) medición del evento*

Se obtuvo información sobre la frecuencia y duración de los episodios de diarrea semanalmente en la casa de cada uno de los niños. Un médico o una enfermera realizó la visita, entrevistó a la madre y examinó al niño. Un episodio de diarrea se definió como tres o más evacuaciones acuosas en un día o cualquier evacuación acuosa con moco o sangre. El episodio se dio por terminado cuando los síntomas no se presentaron por lo menos en dos días consecutivos. Cuando no estaba la madre o algún informante, se hicieron hasta tres visitas por semana. Únicamente las semanas en las que se realizó la visita contribuyeron para el tiempo de seguimiento.

Adicionalmente, se obtuvo información sobre características sanitarias, demográficas y nivel socioeconómico hacia el final del estudio. Estos datos sólo se obtuvieron para 250 niños.

*d) análisis*

La incidencia de diarrea se registró semanalmente. Cada niño contribuyó con un máximo de 12 semanas de seguimiento durante los periodos de tres meses entre cada una de las evaluaciones del estado nutricional. Para cada una de las categorías de estado nutricional, se obtuvieron los años-persona con los que cada niño contribuyó a cada categoría.

## RESULTADOS

A continuación se presenta un cuadro en el que se indica i) la tasa de diarrea por años-persona ("episodios/año"), y ii) los años-persona tanto en niños como en niñas.



**Cuadro I Número de episodios de diarrea por año-persona observación, por grupo de edad y sexo. Datos basales de niños menores de dos años, Tlalpan, México, 1984**

| | *Niños* | | *Niñas* | |
|---|---|---|---|---|
| *Edad en meses* | *Años persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** | *Años persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** |
| 0-5 | 6 | 5.8 | 6 | 7.9 |
| 6-11 | 25 | 5.8 | 24 | 4.9 |
| 12-17 | 42 | 4.6 | 39 | 3.5 |
| 18-23 | 42 | 4.3 | 36 | 2.7 |
| 24-29 | 23 | 2.2 | 21 | 2 |
| 30-36 | 3 | 2.1 | 6 | 1 |

* APO$^{-1}$ = 1/años-persona observación

A continuación se muestra el extracto de un cuadro con las tasas de incidencia de diarrea de los tres grupos de estado nutricional.

**Cuadro II** **Riesgo relativo de diarrea de acuerdo con las diferentes clasificaciones de estado nutricional. Datos basales de niños menores de dos años, Tlalpan, México, 1984**

| *Estado nutricional* | *Años-persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** | *Razón de tasas cruda (IC95%)* |
|---|---|---|---|
| Peso para la edad | | | |
| Normal | 101 | 3.3 | |
| Levemente bajo | 135 | 3.7 | |
| Moderadamente bajo | 37 | 6.0 | |

* APO$^{-1}$ = 1/años-persona observación

## PREGUNTAS

1. Tomando en cuenta la información de la introducción, ¿qué objetivo e hipótesis se plantearía usted?
2. ¿Qué tipo de estudio epidemiológico recomienda para resolver esta pregunta de investigación? Comente.
3. La característica de los estudios de cohorte es que los sujetos se eligen de acuerdo con la exposición de interés. En este estudio, ¿cuál fue la exposición y cuál el evento de interés?
4. En el contexto del presente estudio, mencione una diferencia y una similitud entre los estudios de cohorte prospectivos y los ensayos clínicos. Comente qué implicaciones éticas tiene esta diferencia.
5. Considerando la información resumida en el análisis, ¿qué medida de efecto utilizaría para evaluar la relación entre el nivel nutricional y la presencia de diarrea?
6. Con los datos del cuadro I, calcule el número de casos en:
   - niños menores de 18 meses,
   - niños de 18 meses o más,
   - niñas menores de 18 meses y
   - niñas de 18 meses o más.

   (Señale la fórmula usada para calcular los casos).

Ahora que ya cuenta con el número de casos nuevos de diarrea y con los tiempos-persona:

7. Calcule la tasa de incidencia de diarrea para los niños menores de 18 meses y para los de 18 meses o más. Compare estos resultados en términos absolutos y relativos. Repita el cálculo para las niñas e interprete sus resultados.
8. Considerando la información del cuadro II, estime las razones de tasas de incidencia de diarrea crudas y los intervalos de confianza, tomando como grupo basal el de estado nutricional normal. Complete el cuadro e interprete sus resultados.
9. Tomando en cuenta los resultados del cuadro I, diga usted ¿por cuáles variables ajustaría en el análisis?

## RESPUESTAS

1. El objetivo planteado por los autores fue probar la hipótesis de que la desnutrición preexistente influye en la ocurrencia de enfermedad diarreica en niños preescolares.
2. Los investigadores optaron por un estudio de cohorte. Este tipo de estudio es adecuado, ya que la exposición precede al evento y permite evaluar causalidad.
3. La desnutrición leve y desnutrición moderada corresponden a la exposición de interés; la diarrea corresponde al evento de interés.
4. En ambos estudios se tienen grupos con diferentes grados de exposición (p. ej. en el presente trabajo, la exposición se categorizó en: sin desnutrición, desnutrición moderada y desnutrición severa) que se siguen en el tiempo hasta que ocurre el evento de interés (diarrea). La diferencia radica en que en los estudios de cohorte prospectiva el investigador observa a los sujetos después de ocurrida la exposición, mientras que en los ensayos clínicos el investigador es quien asigna intencionalmente la exposición. De acuerdo con el principio ético de no maleficencia, un investigador nunca podría inducir intencionalmente alguna forma de desnutrición en una población o sujeto.
5. La densidad de incidencia, o tasa de incidencia, es la medida de efecto más adecuada, dada la información recolectada en el presente estudio.
6. El número de casos puede obtenerse a partir de la fórmula de la tasa de incidencia = número de casos / tiempo-persona.



En la página siguiente se presenta el cuadro con el número de casos por grupo de edad (cuadro III).

7. a) Niños:
   Tasa de incidencia (TI) en menores de 18 meses:

$$TI = 373/73 = 5.11 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: se observaron 5.11 casos de diarrea por año-persona en el grupo de niños menores de 18 meses de edad.*

Tasa de incidencia (TI) en niños con edad igual o mayor a 18 meses:

$$TI = 238/68 = 3.5 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: se observaron 3.5 casos de diarrea por año-persona en el grupo de niños de 18 meses o más de edad.*

Comparación en términos absolutos (diferencia de tasas de incidencia, DTI).

$$DTI = 5.11\text{-}3.5\text{=}1.61 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: en el grupo de niños menores de 18 meses se observaron 1.61 más casos por año-persona que en el grupo de niños de 18 meses o más de edad.*

**Cuadro III** **Número de episodios de diarrea por año-persona observación, por grupo de edad y sexo. Datos basales de niños menores de dos años, Tlalpan, México, 1984**

| | *Niños* | | | *Niñas* | | |
|---|---|---|---|---|---|---|
| *Edad en meses* | *Número de casos* | *Años persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** | *Número de casos* | *Años persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** |
| 0-5 | 35 | 6 | 5.8 | 47 | 6 | 7.9 |
| 6-11 | 145 | 25 | 5.8 | 118 | 24 | 4.9 |
| 12-17 | 193 | 42 | 4.6 | 137 | 39 | 3.5 |
| Subtotal | 373 | 73 | | 302 | 69 | |
| 18-23 | 181 | 42 | 4.3 | 97 | 36 | 2.7 |
| 24-29 | 51 | 23 | 2.2 | 42 | 21 | 2 |
| 30-36 | 6 | 3 | 2.1 | 6 | 6 | 1 |
| Subtotal | 238 | 68 | | 145 | 63 | |
| Total | 611 | 141 | | 447 | 132 | |

* APO-1 = 1/años-persona observación.

Número de casos en:
- niños menores de 18 meses = 373
- niños de 18 meses o más = 238
- niñas menores de 18 meses = 302
- niñas de 18 meses o más = 145



Comparación en términos relativos (razón de tasa de incidencia, RTI).

$$RTI = 5.11/3.5 = 1.46$$

*Interpretación: el grupo de niños menores de 18 meses tuvo 1.46 veces el riesgo de presentar diarrea en comparación con el grupo de niños de 18 meses o más de edad.*

*El grupo de niños menores de 18 meses tuvo 46% más riesgo de presentar diarrea que el grupo de niños de 18 meses o más de edad.*

b) Niñas:

Tasa de incidencia (TI) para menores de 18 meses:

$$TI = 302/69 = 4.38 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: se observaron 4.38 casos de diarrea por año-persona observación en el grupo de niñas menores de 18 meses de edad.*

Tasa de incidencia para niñas con edad igual o mayor a 18 meses:

$$TI = 145/63 = 2.30 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: se observaron 2.30 casos de diarrea por año-persona en el grupo de niñas de 18 meses o más de edad.*

Comparación en términos absolutos

$$DTI = 4.38\text{-}2.30 = 2.08 \text{ casos} \times \text{apo}^{-1}$$

*Interpretación: en el grupo de niñas menores de 18 meses se observaron 2.08 más casos por año-persona que en el grupo de niñas de 18 meses o más de edad.*

Comparación en términos relativos

$$RTI = 4.38/2.30 = 1.90$$

*Interpretación: el grupo de niñas menores de 18 meses tuvo 1.90 veces el riesgo de presentar diarrea en relación con el grupo de niñas de 18 meses o más de edad.*

*El grupo de niñas menores de 18 meses tuvo 90% más riesgo de presentar diarrea que el grupo de niñas de 18 meses o más de edad.*

Asimismo, podría usted hacer comparaciones por sexo.

8. A continuación se presenta el cuadro II, ahora con las razones de tasas e intervalos de confianza (cuadro IV). Posteriormente, dos maneras de interpretar los resultados.



**Cuadro IV Riesgo relativo de diarrea de acuerdo con las diferentes clasificaciones de estado nutricional. Datos basales de niños menores de 2 años, Tlalpan, México, 1984**

| *Estado nutricional* | *Años-persona* | *Tasa de incidencia por 1 APO$^{-1}$ ** | *Razón de tasas cruda (IC95%)* |
|---|---|---|---|
| Peso para la edad | | | |
| Normal | 101 | 3.3 | 1.0 |
| Levemente bajo | 135 | 3.7 | 1.1 (0.96-1.26) |
| Moderadamente bajo | 37 | 6.0 | 1.8 (1.52-2.13) |

* APO$^{-1}$ = 1/años-persona observación

Cálculo de intervalos de confianza:

A partir de los datos del cuadro se calcula el número de casos expuestos y el de no expuestos.

$$\text{IC } 95\% = \exp \ln (1.1) \pm 1.96 * \sqrt{1/500 + 1/333}$$

$$\text{IC } 95\% = \exp \ln (1.8) \pm 1.96 * \sqrt{1/222 + 1/333}$$

*Interpretación: los niños en el grupo de desnutrición leve tuvieron 1.1 veces el riesgo de presentar episodios de diarrea en comparación con los del grupo con nutrición normal. Sin embargo, la relación no fue estadísticamente significativa.*

*Interpretación: los niños en el grupo de desnutrición moderada tuvieron 80% más riesgo de presentar episodios de diarrea que el grupo con nutrición normal. Esta asociación fue estadísticamente significativa.*

9. Tanto la edad como el sexo indican estar asociados a la exposición; si estas variables además resultan estar asociadas con el evento de interés, deberían considerarse en el análisis multivariado como potenciales confusores.

# VII Estudios de casos y controles

*Eduardo Lazcano Ponce*
*Eduardo Salazar Martínez*
*Mauricio Hernández Ávila*

"*El objetivo principal de un estudio de casos y controles es proveer una estimación válida y razonablemente precisa, de la fuerza de asociación de una relación hipotética causa-efecto*".

Philip Cole[1]

Es difícil identificar claramente el origen de los estudios de casos y controles en la literatura médica. Según Paneth,[2] los primeros reportes que podrían considerarse como precursores de los estudios de casos y controles identificados en la literatura anglosajona corresponden al francés Pierre Charles Alexander Louis, quien en 1844 utilizó una forma parecida a este tipo de diseño para estudiar la tuberculosis; al inglés John Snow, quien en 1855 investigó el origen de la epidemia de cólera en Inglaterra comparando las caraterísticas de los enfermos con las características de quienes no habían desarrollado la enfermedad;[3] y a Paget y Baker, quienes en 1862 presentaron un análisis sobre la relación entre estado marital y cáncer de mama, comparando mujeres con cáncer mamario con mujeres con otros tipos de cáncer.[4] La primera investigación que utilizó la concepción moderna de los estudios de casos y controles se atribuye a Lane-Claypon, quien en 1926 publicó un reporte sobre factores reproductivos y cáncer de mama basado en el estudio de 500 pacientes hospitalizados y 500 controles.[4] Sin embargo, es hasta los años cincuenta cuando se identifica este tipo de estudio como un diseño epidemiológico específico, en los trabajos reportados por Cornfield[5] y Mantel y Haenszel.[6] Estos autores proporcionaron las primeras bases metodológicas y estadísticas para su aplicación y análisis. Finalmente, en los años setenta, Miettinen[7] establece la concepción contemporánea de este tipo de estudios, al presentar las bases teóricas que establecen la estrecha relación que existe entre este diseño y los estudios tradicionales de cohorte. En México, a pesar de que se habían usado previamente al inicio de la década de los setenta, la concepción y utilización teórica rigurosa de este tipo de estudio comenzó en la década de los ochenta. Uno de los primeros ejemplos es el estudio de Aznar-Ramos y colaboradores,[8] quienes evaluaron la asociación entre enfermedad cardiovascular no reumática y consumo de anticonceptivos orales mediante un estudio de casos y controles pareado por edad, nivel de escolaridad y paridad.

Con estos antecedentes, es posible afirmar que la información derivada de diferentes estudios de casos y controles ha sido notoriamente útil para modificar políticas de salud

y avanzar en el conocimiento médico. A este respecto, estos estudios se han empleado exitosamente para poner en evidencia la asociación entre consumo de cigarrillos y el riesgo de contraer cáncer de pulmón;[9] la exposición al asbesto y la elevada frecuencia de mesoteliomas;[9] y el consumo de estrógenos (dietilestilbestrol) durante el primer trimestre del embarazo y el cáncer de vagina.[10] Si bien podría pensarse que el diseño de cohorte conjunta los factores idóneos para la observación epidemiológica, su realización está seriamente limitada por la ausencia de poblaciones especiales en quienes construirlo y, frecuentemente, por la carencia de tiempo o de los recursos financieros necesarios para observar los grandes grupos poblacionales que se requieren para el análisis de enfermedades poco frecuentes. Por esta razón, los estudios de casos y controles constituyen una alternativa costo-efectiva para identificar factores de riesgo y generar hipótesis que puedan posteriormente ser contrastadas con diseños más robustos desde el punto de vista de la inferencia causal. Los estudios de casos y controles tienen diversas ventajas y desventajas que se resumen en el cuadro I.

**Cuadro I Ventajas y desventajas de los estudios casos y controles**

*Ventajas*

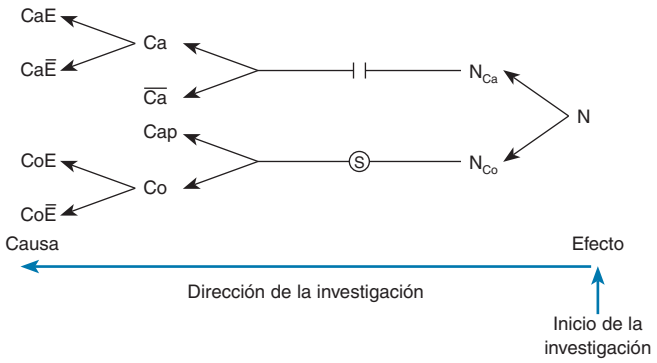
1. Útiles para estudiar problemas de salud poco frecuentes
2. Indicados para el estudio de enfermedades con un largo periodo de latencia
3. Suelen exigir menos tiempo y ser menos costosos que los estudios de cohorte
4. Caracterizan simultáneamente los efectos de una variedad de posibles factores de riesgo del problema de salud que se estudia
5. No es necesario esperar mucho tiempo para conocer la respuesta
6. Requiere de menor número de sujetos en quienes se puede profundizar
7. Estima cercanamente el riesgo relativo verdadero, si se cumplen los principios de representatividad, simultaneidad y homogeneidad

*Desventajas*

1. Especialmente susceptible a sesgos porque:
   - La población en riesgo a menudo no está definida (a diferencia de los estudios de cohorte)
   - Los casos seleccionados por el investigador se obtienen a partir de una reserva disponible
   - Es difícil asegurar la comparabilidad de factores de riesgo poco frecuentes
   - Pueden generar frecuentemente sesgos de información, debido a que la exposición —en la mayoría de los casos— se mide, se reconstruye o se cuantifica, después del desarrollo de la enfermedad
   - Se puede introducir un sesgo de selección, si la exposición de interés determina diferencialmente la selección de los casos y los controles
2. El riesgo o la incidencia de la enfermedad no se puede medir directamente, porque los grupos están determinados no por su naturaleza sino por los criterios de selección de los investigadores
3. Si el problema de salud en estudio tiene una prevalencia mayor de 5%, la razón de momios no ofrece una estimación confiable del riesgo relativo
4. No sirven para determinar otros posibles efectos de una exposición sobre la salud, porque se ocupan de un solo resultado
5. Inapropiados cuando la enfermedad bajo estudio se mide en forma continua

Los estudios de casos y controles representan una estrategia muestral en la que, de manera característica, se selecciona la población en estudio con base en la presencia (caso) o ausencia (control o referente) del evento de interés. Es común que para su realización se utilicen sistemas de registro de eventos relacionados con la salud y padecimientos, por ejemplo, listados de pacientes hospitalizados o reportes de casos a un registro de cáncer, entre otros; para identificar y seleccionar de manera costo-efectiva a grupo de casos; asimismo, que una vez delimitada la población fuente —definida como aquella de donde se originan los casos—, se utilice esta misma para la selección de los controles, los cuales deberán representar de manera adecuada a los miembros de la población fuente que no desarrollaron el evento en estudio. Esto último se cumple siempre y cuando, en el supuesto teórico de haber desarrollado el evento, hubiesen sido identificados como casos en nuestro estudio. Una vez seleccionados los casos y los controles, se compara la exposición relativa de cada grupo con diferentes variables o características que pueden tener relevancia para el desarrollo de la condición o enfermedad.

En teoría, los estudios de casos y controles se basan en la identificación de los casos incidentes en una determinada población durante un periodo de observación definido, tal y como se lleva a cabo en los estudios de cohorte. La diferencia estriba en que, en el estudio de casos y controles, se identifica la cohorte, se identifican los casos y se obtiene una muestra representativa de los individuos en la cohorte que no desarrollaron el evento en estudio. A pesar de que no es necesario conocer el tamaño de la población en riesgo, esto último tiene el propósito de estimar la proporción de individuos expuestos y no expuestos en la cohorte o población base, y evitar, de esta manera, determinar la presencia de la exposición en todos los miembros de la población o cohorte en estudio, lo cual representa un ahorro de recursos y tiempo.[11] En este sentido, la principal diferencia entre los diseños de cohorte y de casos y controles se encuentra en la selección de los sujetos. En un estudio de cohorte se conoce con precisión el tamaño de los grupos en riesgo que son seleccionados a partir de la exposición; es decir, se parte de un grupo de individuos inicialmente exentos de la enfermedad o evento de estudio, y se les sigue en el tiempo con el fin de registrar la ocurrencia del evento. En contraste, en el estudio de casos y controles tradicional, se seleccionan los sujetos en función de la presencia (casos) o ausencia de la enfermedad (controles); y posteriormente, se estiman en ambos grupos las diferencias en la exposición (figura 1). Esto es lo que constituye el paradigma de este tipo de diseños, y repercute ampliamente en su interpretación, aplicación y principales limitantes. En los estudios de cohorte se comparan dos o más grupos de exposición, todos los sujetos de estudio están libres de la enfermedad y en riesgo de desarrollarla al inicio del estudio y se estima la posibilidad o riesgo de presentar el evento o enfermedad a lo largo del tiempo y se relaciona ésta con la condición o exposición estudiada. Los estudios de cohorte tienen una relación causal correcta, se parte de la causa y se observa el efecto. Los estudios de casos y controles, por el contrario, inician con un grupo de sujetos que ya tienen la enfermedad o evento en estudio, parten del efecto en busca de la posible causa, por lo que se considera que no cuentan con una relación de causa-efecto correcta, es decir, no se puede asegurar que la causa precede al efecto y por esto los resultados de este tipo de estudio pueden presentar errores o sesgos. Otra limitación de estos estudios es que, en general, al no conocerse el tamaño de los grupos de riesgo, no pueden estimarse de manera directa las medidas de incidencia que tradicionalmente se obtienen en los estudios de cohorte.




N = población fuente
$N_{Ca}$ = población fuente de los casos
$N_{Co}$ = población fuente de los controles
Ⓢ = muestra de controles
⊣ ⊢ = tamizaje para casos incidentes
Ca = casos de la enfermedad bajo estudio
$\overline{Ca}$ = no casos (pueden ser casos de otras enfermedades)
Cap = casos prevalentes
Co = controles
CaE = casos expuestos
CaĒ = casos no-expuestos
CoE = controles expuestos
CoĒ = controles no-expuestos

**Figura 1 Diseño clásico de un estudio de casos y controles.**

En la mayoría de los estudios de casos y controles sólo es posible estimar *seudotasas*, conocidas como *momios*, mismas que indican la frecuencia relativa de la exposición o condición en estudio entre los casos y controles. La *seudotasa* de exposición en los casos se estima dividiendo los casos expuestos sobre los no expuestos. De manera similar, la *seudotasa* de exposición en los controles se estima dividiendo los controles expuestos entre los no expuestos. El cociente de estas *seudotasas* se conoce como la *razón de momios* (RM) o momios relativos. La RM, bajo ciertas suposiciones que se detallan más adelante, puede ser un estimador no sesgado de la razón de tasas de incidencia o del *riesgo relativo* (RR), las medidas de asociación que tradicionalmente se derivan de los estudios de cohorte y que se utilizan para valorar la fuerza de la asociación entre una exposición y un evento en ese tipo de estudios.

## ESTIMACIÓN DEL RIESGO RELATIVO EN LOS ESTUDIOS DE CASOS Y CONTROLES

Actualmente se reconoce que la tasa de incidencia es la mejor medida de riesgo, esto es, la probabilidad de sufrir un evento; y que la razón de incidencia acumulada es una buena medida para conocer en términos "relativos" lo que aumenta o disminuye dicho riesgo en presencia o ausencia de cierta exposición o condición. Sin embargo, con fines de reconocimiento de causas, resulta de suma importancia práctica la estimación de la razón de momios que se obtiene en los estudios de casos y controles, que en circunstancias específicas se considera como un buen estimador del RR. Los estudios de casos y controles pueden conceptualizarse como una estrategia metodológica para estudiar una cohorte. Dentro de este contexto, pueden existir diferentes alternativas para la selección de los casos. La situación ideal es la que más se acerca a la cohorte, siguiendo el paradigma de los estudios longitudinales; en este sentido, se recomienda seleccionar los casos conforme se diagnostican y aparecen en el sistema de registro utilizado; es decir, la población de casos queda compuesta principalmente por casos incidentes o de diagnóstico reciente. Así, en el otro extremo y que podría considerarse como menos recomendable está el estudiar a los casos existentes en un punto en el tiempo, es decir, los casos prevalentes, que representan una mezcla de casos de reciente diagnóstico y de sobrevivientes o casos que no han emigrado fuera de la población de estudio. A continuación se describen las diferentes opciones para la selección de los casos y los controles.



## MÉTODOS PARA LA SELECCIÓN DE CASOS Y CONTROLES

### Selección de los casos

Considerando que los diseños de casos y controles deben plantearse en función de una cohorte hipotética y que existen dos tipos generales de estudios de cohorte, hay, en consecuencia, dos modalidades de selección de casos: 1) los que semejan las cohortes de densidad de incidencia donde los casos se reclutan conforme se hacen incidentes, y 2) los de incidencia acumulada, donde los casos se seleccionan al final del estudio, siempre definiendo un periodo de tiempo como criterio de elegibilidad. Posiblemente se podría considerar un tercer grupo en el que se seleccionan los casos existentes (sin considerar el tiempo de diagnóstico). Sin embargo, en esta última estrategia estaríamos más cerca de un estudio de prevalencia, por lo que no será considerada esta opción.

En los estudios de casos y controles de incidencia acumulada, el supuesto que debe cumplirse es que la enfermedad es rara, por lo que la prevalencia y la incidencia son pequeñas, entonces la RM es un estimador válido del RR (riesgo relativo), pero en general la RM tiende a sobreestimar el RR cuando ambos son $>1$ o a subestimarlo cuando RR y RM es $<1$. Cuando el supuesto de que la enfermedad es rara no se cumple (como en muchos estudios de brote), entonces la RM está sesgada y hay que corregir; para esto se recomienda utilizar, entre otros, un método propuesto por Greenland para aproximación de varianzas cuyos resultados permiten la estimación de intervalos, por lo que puede ser explorada en forma más precisa la magnitud del efecto.[12]

En el diseño que utiliza el muestreo de densidad de incidencia, donde los casos se reclu-

tan conforme ocurren y los controles se obtienen del mismo grupo poblacional de riesgo, no se necesita el supuesto de enfermedad con una prevalencia menor de 5% y, en este caso en particular, con la RM se puede estimar directamente la razón de tasas de incidencia.

En este contexto tenemos dos modalidades de selección de casos:

1. *Utilización de casos incidentes con periodos de exposición o latencia prolongados*. Esta opción correspondería al posible abordaje de una cohorte de densidad de incidencia, donde la RM tiende a constituirse como un buen estimador de la razón de tasas de incidencia cuando los casos del estudio son incidentes, es decir, se seleccionan a lo largo del estudio conforme ocurren y la exposición que les precede es de larga o corta duración. Este tipo de casos tiene tres ventajas en comparación con los casos prevalentes: a) puede recordarse mejor la experiencia pasada por ser más reciente. Aunque es necesario aclarar que en estudios de casos y controles, el sesgo de memoria no sólo se presenta cuando los casos puedan recordar mejor la exposición, sino que puedan recordarla mejor a diferencia de los controles; b) la supervivencia no está condicionada por los factores de riesgo, como podría ocurrir en los casos prevalentes, y c) es menos susceptible al sesgo de causalidad reversa, y es poco probable que el estatus de enfermedad pueda modificar la exposición que se está estudiando; por ejemplo, la asociación entre prematurez y tabaquismo durante el embarazo[13] o la infección por virus de papiloma humano (VPH) y cáncer cervical.[14]
2. *Utilización de casos prevalentes con periodos de exposición prolongados (muestreo basado en incidencia acumulada).* La RM se parece al riesgo relativo (RR) si, a pesar de utilizar casos prevalentes, el periodo de exposición es muy largo y la enfermedad no afecta el estado de exposición y la enfermedad es rara. Los casos prevalentes que ocurren en cierto periodo de tiempo pueden incluirse, especialmente, cuando no se dispone de casos nuevos porque la enfermedad es muy rara y tiene baja letalidad, y cuando la exposición no modifica el curso clínico (sobrevida) de la enfermedad, como es el caso de enfermedades de predisposición genética. Ejemplo: el estudio de una proteína nuclear específica cerebral asociada al riesgo de esquizofrenia.[15]



**Selección de los controles**

El grupo control o referente, en los estudios de casos y controles, se utiliza fundamentalmente para estimar la proporción de individuos expuestos y no expuestos en la población base (fuente) de la cual se muestrearon los casos. Por esta razón, la fuente poblacional de los controles quedará definida en la medida en que se expliciten claramente los criterios de selección que se utilizaron para conformar el grupo de casos, así como la población hipotética de origen. Como regla general se puede decir que el grupo control más apropiado corresponde a los individuos de la misma población que dio origen a los casos, que están en riesgo de desarrollar el evento en estudio y que en el caso hipotético de que lo desarrollaran aparecerían en el estudio como casos.

Cuando los casos se obtienen de una población claramente definida en tiempo, espacio y lugar, y constituyen un censo de los eventos en estudio o una muestra representativa de los mismos, la selección de controles puede realizarse mediante un muestreo aleatorio simple de la población base en riesgo, siguiendo el esquema de selección de los casos, ya sea de densidad de incidencia o de incidencia acu-

mulada. En este caso, la selección es un simple procedimiento técnico que no introduce ningún sesgo más allá de errores muestrales que podrían existir al utilizar la totalidad de la base poblacional como marco muestral para la selección de los controles. Como ejemplo de esta situación, podemos citar el estudio realizado en la Ciudad de México por Romieu y colaboradores.[16] Los autores conformaron el grupo de casos con una muestra representativa de casos incidentes de cáncer de mama que ocurrieron entre los habitantes de la Ciudad de México entre 1990 y 1992; la selección de los casos en el estudio antes mencionado siguió un esquema de densidad de incidencia. Para conformar el grupo control los autores seleccionaron una muestra representativa de las mujeres residentes en esa ciudad durante el mismo periodo y del mismo grupo de edad. La selección del grupo control se llevó a cabo durante el mismo tiempo en que se seleccionaron los casos, por lo que se puede suponer que los controles representan la población que dio origen a los casos en el momento en que se registró cada caso. La suposición en lo que se refiere a este grupo es que, en caso de que se les hubiera diagnosticado cáncer de mama a las mujeres seleccionadas, habrían sido detectadas y se habrían incluido como casos; de hecho bajo este diseño una misma participante puede ser seleccionada inicialmente como control y posteriormente como caso, en el supuesto que desarrollara cáncer de mama durante el periodo de estudio. Bajo este diseño muestral podemos entonces afirmar que el grupo control seleccionado representa adecuadamente a la población base durante el periodo de estudio, y puede utilizarse para hacer inferencias válidas sobre la proporción relativa de sujetos expuestos y no expuestos en la población base que dio origen a los casos.

En otro estudio, también realizado en la Ciudad de México, Pérez-Padilla y colaboradores[17] reportaron resultados sobre los factores de riesgo para enfermedad pulmonar obstructiva crónica (EPOC). Para este estudio se seleccionaron como casos las mujeres con diagnóstico reciente de EPOC que acudían para atención médica al Instituto Nacional de Enfermedades Respiratorias (INER). Puesto que el INER es un centro nacional de referencia para padecimientos pulmonares en México, que atiende a la población no asegurada, los pacientes que acuden a este centro hospitalario en busca de atención no están bien caracterizados, razón por la cual los autores no lograron definir una población base en tiempo, espacio y lugar, que les permitiera realizar directamente la selección de los controles. Para superar este problema, los autores eligieron como controles a personas que acuden al INER por otros padecimientos no relacionados con la EPOC. En este caso, la suposición de que el grupo control representa adecuadamente la base de donde se originaron los casos de EPOC se cumple, siempre y cuando pueda suponerse que, en caso hipotético de haber desarrollado EPOC, en lugar del padecimiento que originalmente los llevó al INER, los controles también habrían acudido al INER y habrían sido incluidos en la lista de casos. En la medida en que esta suposición se cumpla, el grupo control será adecuado para estimar la proporción de expuestos y no expuestos en la población base que dio origen a los casos. Por lo tanto, las consideraciones fundamentales para la selección de los controles incluyen:

1. Los controles deben seleccionarse de la misma base poblacional (de la cohorte hipotética) de donde se originaron los casos. Operacionalmente, este concepto quiere decir que, en el supuesto de que el control desarrollara la enfermedad o evento en estudio, necesariamente tendría que aparecer en la lista de casos. Esto último también implica que los controles estuvieron en riesgo de desa-

rrollar el evento al mismo tiempo que los casos.
2. Los controles deben seleccionarse independientemente de su condición de expuestos o no expuestos para garantizar que representen adecuadamente a la población base. Esto último se logra siempre y cuando la condición de exposición no determine la posibilidad de que un individuo sea o no incluido en el estudio como control, lo que implica que las fracciones muestrales para los controles expuestos y no expuestos deben ser las mismas, aunque la mayor parte de las veces son desconocidas por el investigador. En la medida en que éstas difieran, se introducirá un sesgo de selección y se comprometerá la validez interna del estudio.
3. La probabilidad de selección para los controles debe ser proporcional al tiempo que el sujeto estuvo en riesgo de desarrollar el evento o enfermedad en estudio. Así, un individuo que migró o falleció durante la investigación dejará de ser elegible como control en el momento en que sale del estudio. Una manera de operacionalizar este concepto es seleccionando un control del grupo de individuos elegibles cada vez que se detecta o selecciona un caso; lo que se conoce como selección por grupo en riesgo o de densidad de incidencia. En teoría, utilizando este esquema de selección, se asegura que los controles están en riesgo de desarrollar el evento en el momento de la selección. Este esquema también implica que un sujeto elegido como control en una etapa temprana del estudio, también podría seleccionarse como caso en etapas posteriores. En el caso de muestreo por incidencia acumulada, se debe definir claramente el periodo de riesgo y los controles se seleccionan de entre los miembros de la población base que no desarrollaron el evento al terminar el periodo de seguimiento.
4. En la selección de los controles deben evitarse, en la medida de lo posible, los factores de confusión. Se espera que el grupo control sea similar al grupo de casos en lo que se refiere a otras variables que podrían ser factores de riesgo para el desarrollo del evento y, al mismo tiempo, estar asociados con la exposición. Una estrategia frecuentemente utilizada para lograr este requisito es el pareamiento o igualación de atributos. Este esquema implica un segundo requisito de elegibilidad para el control. Por ejemplo, si al escogerlo se decidiera homogeneizarlo (parearlo) con el caso por edad y género, el control debería ser seleccionado al azar del grupo de controles potenciales que son del mismo género y grupo de edad.
5. La medición de variables debe ser comparable entre los casos y los controles. Todos los procedimientos para medir la exposición, los factores de confusión o de modificación de efecto potenciales deben aplicarse, notificarse y registrarse de la misma manera tanto en casos como en controles. Esto último tiende a minimizar la posibilidad de sesgos de información.



En este contexto, existen diferentes posibilidades y fuentes de obtención de controles, a saber:

1. *Con base poblacional.* Si los casos representan una muestra de todos los casos que ocurren en una población identificada y definida claramente en tiempo y espacio, y los controles se muestrean directamente de esta misma población, los controles se definen como base po-

blacional. Este tipo de controles son más factibles de utilizar cuando los casos se identifican mediante registros poblacionales o se cuenta con los suficientes recursos para obtenerlos directamente, como se ejemplifica en el cuadro II[18] y figura 2.

2. *Controles vecindarios.* Este tipo de controles se utiliza frecuentemente cuando no se cuenta con una base poblacional bien definida. El procedimiento de selección consiste en obtener uno o más controles, seleccionados al azar en la misma zona residencial donde habita el caso. Este tipo de selección tiende a homogenizar los casos y controles en lo que corresponde al nivel socioeconómico, a las exposiciones ambientales que ocurren en la zona de residencia y a otro tipo de variables que pudieran estar asociadas con el lugar de residencia. Este procedimiento muestral puede generar sesgos de selección, ya que no siempre puede asumirse que los controles vecinales representan de manera válida la base poblacional de donde se originaron los casos. La suposición de que representan adecuadamente la base poblacional se cumple en el sentido en que si los controles hubiesen desarrollado el evento en estudio entonces hubieran aparecido en el estudio como casos. Tal situación es la planteada por Carrique-Mas y colaboradores, quienes estudiaron los factores de riesgo de campilobacteriosis entre niños de Suecia. Ellos utilizaron 126 casos reportados al sistema nacional de vigilancia y seleccio-

**Cuadro II Ejemplo de casos y controles con base poblacional**

*Objetivo*

Especificar cuál es el evento de interés en el ejemplo mostrado, el cual no muestra la calidad del Programa de Detección Oportuna de Cáncer Cervical (PDOC), sino la asociación entre dicho programa (probablemente su utilización, ya que no es clara esta exposición) y cáncer cervical

*Material y métodos*

Casos y controles con base poblacional. Selección de 513 casos de cáncer cervical de ocho hospitales del área metropolitana de la Ciudad de México y 1 007 mujeres seleccionadas de un muestreo aleatorio de 6 220 viviendas de la Ciudad de México

*En teoría:*

- La muestra de estudio se obtuvo a partir de la población de la que se obtienen todos los casos incidentes
- Los controles se eligieron aleatoriamente entre los miembros de la misma población sin la enfermedad
- Aun cuando los casos se identificaron en ocho hospitales es razonable suponer que representan los casos en el área geopolítica de estudio

*Resultados*

El PDOC en la Ciudad de México no se asocia con cáncer cervical (RM = 0.95; IC 95% 0.76-1.19).

El PDOC (o su utilización, si es el caso) en la Ciudad de México disminuye la posibilidad de presentar cáncer cervical (RM= 0.51; IC 95% 0.39-0.67)

Fuente: referencia 18

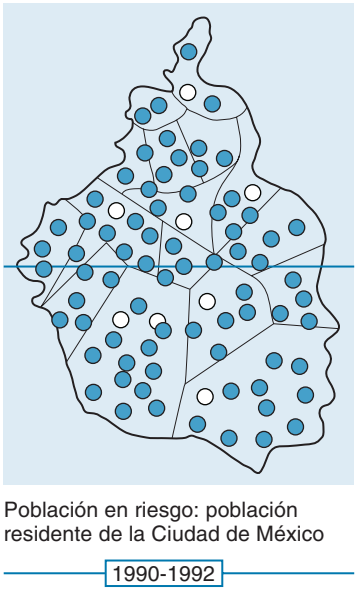



**Figura 2** **Estudio de casos y controles**



naron controles de un registro poblacional pareado por lugar de residencia, entre otros factores.[19]

3. *Controles hospitalarios.* Este grupo control se selecciona al azar de los pacientes que acuden al mismo hospital donde se realizó la selección de los casos, pero que acuden por un padecimiento diferente. La principal limitación de este grupo control es que puede no representar adecuadamente la exposición en la población base de donde se originaron los casos. Esto último se debe a que puede existir una relación entre la exposición en estudio y la causa de demanda de atención; lo cual condicionaría una sobreestimación de la exposición en este grupo control y por lo tanto no representaría adecuadamente a la base poblacional que se encuentra fuera del hospital. Por ello, cuando se utiliza este tipo de controles, los investigadores deben estar seguros de que la exposición en estudio no está relacionada con la enfermedad o el motivo de hospitalización. En consecuencia, es recomendable utilizar una variedad de diagnósticos posibles con el fin de atenuar los efectos de incluir un grupo diagnóstico específico que pudiera estar relacionado con la exposición en estudio. En otras palabras, es probable que al padecer algún tipo de enfermedad difieran de los individuos sanos en una serie de factores que tienen relación con el proceso de enfermar, como pueden ser mayor prevalencia en el consumo de tabaco, alcohol o deficientes hábitos dietéticos, y estos fac-

tores pueden estar relacionados directa o indirectamente con la exposición en estudio. Un ejemplo de la utilización de controles hospitalarios es el descrito por Gelatti y colaboradores, quienes evaluaron la asociación entre consumo de café y el riesgo de cáncer hepatocelular. Ellos reclutaron 250 casos y 500 controles de sujetos hospitalizados por otras razones que neoplasias o enfermedades hepáticas, u otras directamente relacionadas con consumo de alcohol.[20]

4. *Controles seleccionados aleatoriamente de números telefónicos.* Frecuentemente utilizado en países desarrollados, es una selección aleatoria de viviendas del listado telefónico de un área geopolítica. Tiene diversas limitaciones que incluyen la probabilidad de contactar a los sujetos elegibles, porque no siempre están en su casa y varía el número de residentes. Otras limitaciones importantes son: el tiempo que se invierte para contactar a la población objetivo, que puede requerir de varias llamadas adicionales, y la inclusión de números no residenciales que puede afectar la tasa de respuesta; la existencia de varios números telefónicos en una misma casa, y el uso de contestadoras automáticas, que plantea problemas adicionales. La limitante más importante en nuestro país es el gran número de viviendas que no cuentan con teléfono, lo que excluye a una gran cantidad de participantes potenciales. Tal es el caso descrito por Ward y colaboradores en un estudio de factores de riesgo de neuroblastoma donde se habían identificado 101 casos y se realizaron 3 245 llamadas telefónicas, para identificar un control pareado por año de nacimiento y raza. En 25.5% de los números telefónicos rehusaron dar información.[21]
5. *Controles con otras enfermedades de un registro poblacional.* Para la obtención de este tipo de controles pueden utilizarse registros de base poblacional como los de tumores, de sistemas de vigilancia epidemiológica o estadísticas vitales (fallecidos), hospitales, centros de atención primaria, empresas y compañías de seguros, entre otros. La ventaja es la procedencia de una misma base poblacional, pero debe tenerse el cuidado de excluir los diagnósticos que puedan estar relacionados con la exposición bajo estudio. En algunos casos puede constituir una variante de controles hospitalarios o que acuden a una misma práctica médica. Referiremos como ejemplo el estudio de Chiaffarino y colaboradores, quienes realizaron un estudio de 1 031 casos de cáncer de ovario epitelial y seleccionaron 2 411 controles de mujeres admitidas a una misma red hospitalaria con fracturas por condiciones traumáticas, enfermedades ortopédicas no traumáticas, postoperatorio por apendicitis y/o hernia estrangulada, así como otras múltiples causas relacionadas con problemas oculares, óticos, de nariz y dentales.[22]
6. *Controles de amigos o familiares.* Son personas relacionadas con los casos. Este grupo presenta la ventaja de reducir los costos y tienen una elevada probabilidad de que provengan de una misma base poblacional que los casos. Sin embargo, su principal inconveniente es el potencial riesgo de crear un grupo de casos y controles muy homogéneos, dado que es muy común que los amigos compartan estilos de vida, y por esta razón este tipo de controles no representarían de manera adecuada la exposición en la población base. Sin embargo, para trabajos en los que se estudian factores genéticos, pueden constituir un buen grupo

control. Citamos el estudio desarrollado por Cullen y colaboradores donde examinaron la influencia del consumo elevado de edulcorantes en diabéticos dependientes de insulina comparados con un grupo control, que fueron amigos de los pacientes, pareados por edad y sexo.[23]

7. *Controles obtenidos del registro de mortalidad.* En este caso los controles no se seleccionan directamente de la base poblacional —que necesariamente está compuesta por personas en riesgo de padecer el evento—, sino de algún registro de mortalidad proveniente de dicha base. Estos controles pueden ser útiles cuando estamos seguros que en ellos la distribución de la exposición que nos interesa investigar —y que presumiblemente tiene relación con la causa de muerte en los casos— es potencialmente la misma a la observada en la base poblacional. Debe señalarse que este tipo de controles debe restringirse a las categorías de muerte que, en teoría, no están directamente relacionadas con la exposición de interés, pues de otra manera podríamos incurrir en un sesgo de selección, lo que invalidaría la posibilidad de comparar los grupos. Recuérdese que la única diferencia entre los casos y los controles debe ser la presencia o no, respectivamente, del resultado que se desea investigar. Como se señaló con anterioridad, estos controles son útiles cuando se ponen en práctica estudios de mortalidad proporcional. Cerhan y colaboradores[24] describieron un estudio de mortalidad proporcional para comparar los patrones de mortalidad entre agricultores y no agricultores de Iowa, utilizando 88 090 certificados de muerte entre 1987-1993. Tasas de mortalidad proporcional fueron estimadas usando el registro de mortalidad en los no agricultores para generar números esperados.
8. *Uso de controles del mismo tipo o controles de diferentes tipos.* De acuerdo con la disponibilidad de controles es posible que sea costo-efectivo utilizar dos o más controles por cada caso, esto último puede incrementar el poder estadístico del estudio; sin embargo, se acepta que se gana incremento en el poder solamente hasta un índice de cuatro controles por cada uno de los casos.[25]

   También pueden utilizarse varias fuentes de controles (*múltiples controles de diferentes tipos)*. En este caso, puede elegirse un grupo adicional de controles vecindarios o poblacionales, en espera de que los resultados obtenidos cuando los casos se comparan con controles hospitalarios sean similares a los resultados obtenidos cuando se comparan con otro tipo de controles. El problema es que si los hallazgos difieren, la razón de la discrepancia no puede identificarse fácilmente. A final de cuentas, la utilización de varias fuentes de controles se utiliza para evaluar la consistencia de los resultados independientemente del grupo control. Un ejemplo de esta estrategia fue recientemente publicado. Calin y colaboradores evaluaron un polimorfismo en ARLTS1 asociado como predisponente a cáncer familiar, utilizando 216 pacientes con tumores, 109 sujetos con cáncer familiar o múltiples cánceres y 475 sujetos con enfermedades diferentes a cáncer.[26]
9. *Pareamiento.* El pareamiento se desarrolló como una respuesta intuitiva a la falta de aleatorización en los estudios observacionales. La asignación aleatoria de la población de estudio a la exposición (tratamiento) iguala la distribución en los grupos que se comparan. Una de las

limitaciones de los estudios de casos y controles es que los resultados pueden estar distorsionados debido a la falta de comparabilidad de los casos y los controles respecto a variables que están asociadas tanto con la exposición como con el evento en estudio. Una manera de controlar esta fuente de error es utilizando un diseño pareado. Este tipo de estrategia además de controlar el efecto de factores de confusión, aumenta la eficiencia estadística del diseño, lo que se traduce en una disminución del tamaño de muestra necesario para llevar a cabo el estudio. En este tipo de diseño los controles se seleccionan en función del valor que toman las variables potencialmente confusoras en la serie de casos y de esta manera se igualan los grupos. Por ejemplo si el caso seleccionado es del género masculino y del grupo de edad de entre 35 y 40 años, se deberá seleccionar un control del mismo género y del mismo grupo de edad; este procedimiento se repite para cada caso seleccionado y de esta manera se cumple el supuesto de que las variables potencialmente confusoras tomarán el mismo valor promedio en ambas series de casos y controles. Este tipo de estrategia de selección de controles puede asegurar la homogeneidad por edad y sexo, y facilitar la comparación de casos y controles en presencia de exposiciones que varían con el tiempo. En general podemos decir que todos los estudios de casos y controles en los que no se seleccionan los controles de manera aleatoria tienen cierto grado de pareamiento. La selección de controles vecinales es una forma de pareamiento por factores ambientales, clase social o variables no medidas asociadas al lugar de residencia. En general se utilizan dos tipos de pareamiento: individual o grupal (pareamiento por grupos de frecuencia) y éstos pueden ser univariados o multivariados dependiendo de la complejidad del estudio. Este tipo de estrategia de selección de controles tiene varias desventajas: los costos del estudio aumentan ya que por un lado la selección final de controles requiere de información previa a la selección, y de no encontrarse controles adecuados, se tendrían que excluir los casos. El análisis estadístico es más complejo ya que no es posible estimar la razón de momios de manera directa y se requiere de un análisis condicional que incorpora la complejidad del diseño pareado. Esto último permite corregir el sesgo de selección que se introduce al seleccionar los controles en relación a variables que se asocian de manera significativa con la exposición y el evento; este sesgo tiende a subestimar la asociación real. Además no es posible estimar el riesgo de las variables de pareamiento ya que por diseño son invariantes entre casos y controles. Otra posible desventaja es la de sobrepareamiento (éste ocurre cuando el pareamiento se lleva a cabo utilizando una variable que está asociada con la exposición y no con el evento), el efecto que tiene esto es que reduce la eficiencia del estudio. Cuando se parea utilizando variables que pueden ser modificadas por la exposición o el evento (como intensidad de síntomas) se puede condicionar la introducción de un sesgo; igualmente el estudio puede perder validez si se parea por una variable que es una condición intermedia en el camino causal entre la exposición y la enfermedad. Recientemente Dong y colaboradores[27] describieron un estudio de 92 casos de suicidio que fueron pareados en una relación 1 a 1 por las siguientes caracte-

rísticas: a) admisión al mismo hospital, b) sexo, c) edad, d) diagnóstico, e) duración de estancia en un servicio psiquiátrico, f) datos de admisión. Para propósitos del pareamiento, los diagnósticos fueron agrupados en uso de sustancias adictivas, esquizofrenia, enfermedad bipolar, depresión, y otros.

## VARIANTES DEL DISEÑO DE CASOS Y CONTROLES

El objetivo principal al realizar un estudio de casos y controles es el de estimar un riesgo para un factor o exposición determinado que sea lo más cercano al que se habría obtenido en caso de haber podido realizar una cohorte prospectiva. Por esta razón, tanto los casos como los controles, en teoría, deben tener características de representatividad, simultaneidad y homogeneidad.

Representatividad significa que los casos deben representar a todos los casos (eventos) que ocurrieron en la cohorte hipotética, en un tiempo determinado, y que los sujetos que se seleccionan como controles deben representar a los sujetos de la cohorte que están en riesgo de convertirse en casos; además, por diseño ambos deben pertenecer a la misma base poblacional de la cohorte hipotética estudiada.

Simultaneidad significa que los controles deben obtenerse en el mismo tiempo que surgieron los casos. Esto se puede operar, como se mencionó anteriormente, utilizando un muestreo de densidad de incidencia o de incidencia acumulada.

Finalmente, homogeneidad significa que los controles deben obtenerse de la misma cohorte de donde surgieron los casos y ese tipo de selección debe realizarse, por lo tanto, cuando no se utiliza un diseño pareado independiente de la exposición bajo estudio. Estas tres características se describen en el cuadro III, donde se observan las limitaciones que por definición tienen las diversas variantes de diseños de casos y controles, y donde cualquier factor que se aleje de estos principios producirá que la medida de efecto —RM— se sobre o subestime, y afecte la validez del estudio.

Durante los últimos años se ha avanzado considerablemente sobre los aspectos metodológicos de los estudios de casos y controles; en particular se ha conceptuado más claramente la estrecha relación que existe entre estos estudios y los de cohorte, lo que ha permitido el desarrollo de diferentes variantes en los esquemas de selección de casos y controles, esto último con el fin de lograr estimaciones de riesgo válidas y de reducir los costos que implica la realización de los grandes estudios de cohorte. Recientemente se han propuesto y formalizado, desde el punto de vista estadístico, variantes que inciden sobre la selección de los casos o la definición del grupo control; a continuación se describen las más utilizadas:

1. *Estudios caso-cohorte*. En esta variante, la definición de casos y controles se encuentra anidada en una cohorte fija, bien definida en tiempo, espacio y lugar, en la cual existe el interés de estimar la razón de incidencia acumulada y se asume que todos los miembros de la cohorte tendrán el mismo tiempo de seguimiento. Para realizar un estudio de caso-cohorte se requiere llevar a cabo los pasos que se describen en la figura 3. En un primer tiempo, se define la cohorte o población en estudio; en un segundo, se selecciona el grupo control que se utilizará para estimar la proporción de individuos expuestos y no expuestos que se encuentran en riesgo de de-

**Cuadro III Características de los estudios de casos y controles**

| Tipo de estudio de casos y controles | *Representatividad* | | *Simultaneidad* |
|---|---|---|---|
| *Homogeneidad* | *Casos* | *Controles* | |
| *Caso-cohorte* | *Sí* | *Sí* | *Asegurada* |
| Definitiva | | | |
| Caso-caso | Azar | Azar | Asegurada |
| Definitiva | | | |
| Casos y controles anidado en una cohorte | Sí | Sí | Asegurada |
| Definitiva | | | |
| Casos y controles con base poblacional | Azar | Sí | Posible |
| Posible | | | |
| Casos hospitalarios y controles con base poblacional | No | Sí, siempre que se obtengan de un marco muestral de la población de la que surgen los casos | Desconocida |
| Desconocida | | | |
| Casos hospitalarios y controles vecindarios | No | No | Desconocida |
| Desconocida | | | |
| Casos y controles hospitalarios | No | No | Desconocida |
| Desconocida | | | |

sarrollar el evento al inicio de la investigación y, posteriormente, se realiza el seguimiento de la cohorte, con el fin de detectar los eventos (casos incidentes) que se desarrollan a lo largo del tiempo y caracterizarlos en términos de su pertenencia al grupo expuesto o no expuesto.

Es evidente que al usar este tipo de selección, un sujeto inicialmente identificado como control podría desarrollar el evento de interés durante el seguimiento y ser seleccionado como caso. Esta última situación, cuando ocurre con frecuencia, puede convertirse en una limitante importante y comprometer el poder estadístico del estudio. Por esta razón, esta estrategia se recomienda para el análisis de enfermedades poco frecuentes, en cohortes fijas, claramente definidas, donde la determinación de la exposición en todos los miembros de la cohorte resultaría muy costosa, como la que se describe en el cuadro IV.[28]

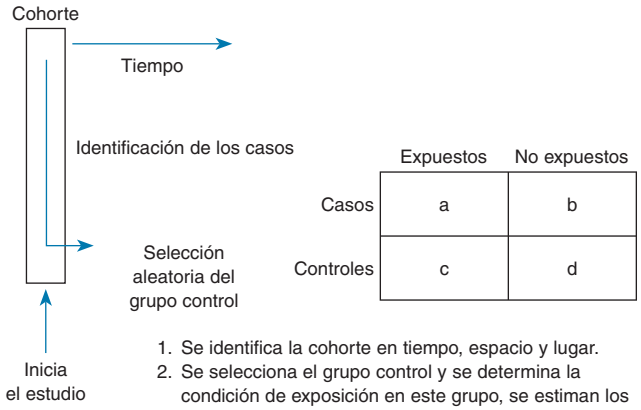



**Figura 3** **Diseño de casos y controles de tipo caso-cohorte**



2. *Estudios de casos y controles anidado o de grupo de riesgo*. En esta variante se utiliza un esquema de muestreo conocido como de grupo de riesgo o de densidad de incidencia, ya que la elegibilidad de un individuo como control depende de que éste se encuentre en riesgo, es decir, que sea miembro de la cohorte y por lo tanto elegible como control en el momento en que se selecciona o identifica un caso. El caso o casos que ocurren en un punto en el tiempo y el conjunto de individuos que se encuentran en riesgo en ese punto constituyen el grupo de riesgo. A lo largo del estudio el grupo de riesgo varía y los controles se seleccionan de forma concomitante con la ocurrencia de los eventos. En esta variante de casos y controles es frecuente asumir que la selección de los casos y controles se lleva a cabo dentro de una cohorte dinámica, donde los sujetos de estudio permanecen durante tiempos variables y en los que la exposición puede cambiar en el tiempo. Para realizar un estudio de casos y controles anidado es necesario llevar a cabo los pasos que se describen en la figura 4. En un primer tiempo se define de manera conceptual la cohorte o población en estudio; se realiza el seguimiento de la misma con el fin de detectar los eventos que ocurren (casos incidentes) a lo largo del tiempo, y cada vez que se identifica un caso, se seleccionan uno o varios controles de la población que en ese momento particular se encontraba en riesgo de desarrollar el evento en estudio. Es evidente que al usar este tipo de selección, un sujeto inicialmente identificado como control podría desarrollar el evento de interés durante el seguimiento

**Cuadro IV Ejemplo de estudio de caso-cohorte**

*Objetivo*

Investigar si las características antropométricas se relacionan con cáncer de próstata (CP)

*Material y métodos*

Estudio de cohorte con una medición basal en 1986 de 58 279 hombres entre 55-69 años. Después de 6.3 años de seguimiento, analizaron 681 casos, de CP y 1 565 miembros de la subcohorte. Se excluyó a hombres con medidas antropométricas incompletas

*Exposición*

Se estimó el índice de composición corporal a la edad de 20 años en relación con la medición basal

*Resultados*

El índice de masa corporal (IMC) no se asoció significativamente con cáncer de próstata. Sujetos con IMC a los 20 años $<19$ fueron categorizados como referencia. Con IMC entre 19-20, la RM fue de 1.06; entre 21-22 fue de 1.09; entre 23-24, de 1.39 y, finalmente, IMC $> 25$ tuvo una RM de 1.33. Esta tendencia fue significativa ($p = 0.02$)

*Conclusiones*

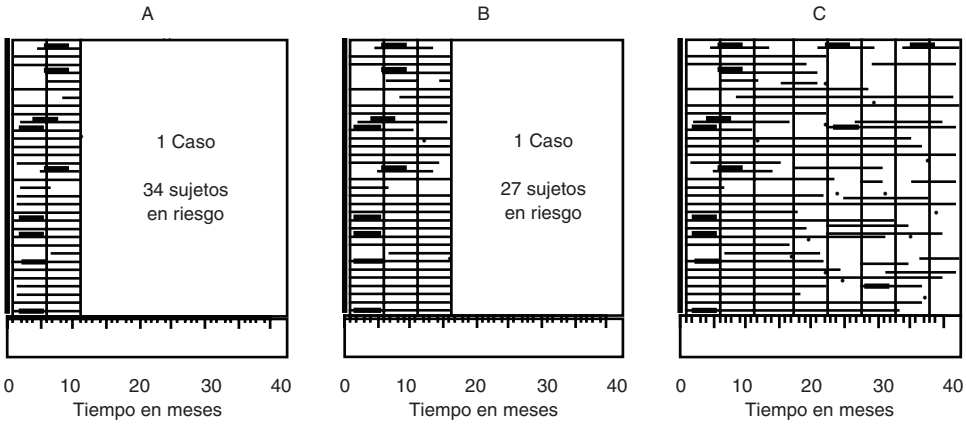
Los hallazgos sugieren que el incremento del IMC a temprana edad puede aumentar el riesgo de desarrollar cáncer de próstata

* Fuente: referencia 28.



y posteriormente ser seleccionado como caso. Esta situación en general ocurre con poca frecuencia, pero depende de la incidencia del evento en estudio. Sin embargo, en ese diseño en particular podría ocurrir que un individuo fuera inicialmente seleccionado como control y posteriormente como caso. Esta situación no es fuente de error o de sesgo, ya que en los estudios de cohorte un mismo individuo puede contribuir tanto al numerador como al denominador y esta misma situación se mantiene en este tipo de estrategia. El diseño de casos y controles anidado, o de grupo de riesgo, se recomienda para el estudio de enfermedades poco frecuentes, en cohortes dinámicas en las que la determinación de la exposición y sus cambios en el tiempo, en todos los miembros de la cohorte, resultaría muy costosa (cuadros V y VI).[29,30]

3. *Estudios caso-autocontrol*. En este tipo de muestreo se utiliza al mismo sujeto que desarrolló el evento como su propio control. Este diseño se ha utilizado para indagar de forma sistemática, si el sujeto de estudio estuvo haciendo algo inusual, o si estuvo expuesto a alguna situación o sustancia poco antes de que se detectara el evento de estudio. Como se mencionará mas adelante, este tipo de estrategia es factible utilizarla para exposiciones que son de corta duración y que cambian en el tiempo y con eventos que son fáciles de detectar. Por ejemplo el uso de teléfono celular durante la conducción de vehículos de automotor y el riesgo de accidentes de tráfico. Para responder la pregunta inicial, se hace la comparación en el mismo individuo en




En los páneles A, B y C se representa gráficamente la realización de un estudio de casos y controles anidado en una cohorte dinámica. En el tiempo 0 se definen los criterios de elegibilidad para la cohorte. En el panel A se observa que el primer caso incidente se detecta a los 10 meses de haber iniciado el estudio; en ese momento en la cohorte existían en riesgo 34 sujetos; este grupo forma la población base para la selección de los controles, es decir, el grupo de riesgo. En el panel B se observa que el segundo caso se detecta a los 16 meses de seguimiento y que el grupo de riesgo ha variado en su composición; en este punto en el tiempo existen 27 sujetos elegibles como controles. Finalmente en el panel C se presenta el resultado final de la definición variable de los grupos de riesgo, que cambia conforme ocurren y se selecciona a los casos.



**Figura 4** **Diseño de casos y controles anidado o de grupo de riesgo.**

diferentes tiempos, esto es, comparar la exposición a ciertos agentes durante el intervalo en que el evento no había ocurrido (periodo de control), con la exposición durante el intervalo en que el evento ocurrió (periodo de caso). Este tipo de diseño puede conceptualizarse como un estudio de casos y controles pareado, en donde cada uno de los individuos sirve como su propio control y donde todos los individuos experimentaron el evento.[31]

Existen tres elementos sustantivos en este diseño: 1) periodo de exposición, 2) el periodo de riesgo, y 3) la presentación del evento de estudio.

En el periodo de exposición se cuantifican los hábitos rutinarios del sujeto de estudio en periodos de tiempo determinados. Eventualmente, pueden existir uno, dos o más periodos de exposición (uno, dos, tres o más grupos considerados autocontrol), que constituyen la fase de alteración de riesgo en una población. El periodo de riesgo, se conceptualiza como la exposición que se presenta justamente antes de presentar el evento o enfermedad bajo estudio. La longitud del periodo de efecto y su periodo de riesgo puede decidirse empíricamente. Sin embargo, este periodo es crítico, porque la sobre o subestimación de la duración puede diluir la posible asociación bajo estudio. Un ejemplo de este tipo de diseño[32] se presenta en el cuadro VII y en la figura 5.

**Cuadro V Ejemplo de casos y controles anidados en una cohorte**

*Objetivo*

Evaluar la posible asociación entre infección por *Helicobacter pylori* y cáncer de estómago en una cohorte de 5 908 hombres japoneses-estadounidenses de Hawaii

*Periodo de estudio*

1967-1970

*Banco de sueros*

Alícuotas obtenidas en cada sujeto al inicio del estudio

*Número de casos reportados en 1989*

109 casos de cáncer gástrico diagnosticados histopatológicamente 20 años después

*Exposición*

Presencia de anticuerpos IgG a *Helicobacter pylori*

*Selección de controles*

Muestreo aleatorio pareado por edad

*Resultados*

94% de prevalencia en casos

76% de prevalencia en controles

RM = 6.0; IC 95% 2.1-17.3

*Conclusiones*

Infección con H. pylori se asocia estrechamente con la aparición de cáncer gástrico

* Fuente: referencia 29

Para el análisis se utiliza el mismo método de un estudio de casos y controles pareado; pero en lugar de caso se utiliza el periodo de riesgo en comparación con el periodo control. Asimismo, los datos también pueden analizarse por medio de las unidades tiempo-persona.

4. *Estudios de mortalidad proporcional.* Tanto los casos como los controles se obtienen de los registros de muerte de una fuente poblacional. Rothman[33] establece que este tipo de controles son aceptables sólo si la distribución de la exposición entre los grupos es similar a la que presenta la base poblacional. Consecuentemente, la serie seleccionada como control deberá restringirse a las categorías de muerte que no están relacionadas con la exposición.
5. *Estudios de caso-caso.* Una estrategia que ha sido utilizada en estudios analíticos de enfermedades infecciosas notificables identificados en sistemas de vigilancia epidemiológica son los estudios de caso-caso. Mediante esta metodología se compara la historia de exposición en subgrupos de casos. El análisis es li-

**Cuadro VI Ejemplo de casos y controles en una cohorte dinámica**

*Objetivo*

Evaluar la relación entre el consumo de estrógenos en la posmenopausia y la incidencia de infarto agudo del miocardio

Periodo de estudio

1978-1984

*Población fuente*

Mujeres entre 50 y 64 años de edad con cobertura de atención médica completa por un grupo de cooperación en salud de Seattle, Washington.

*Casos*

Casos incidentes diagnosticados entre 1978 y 1984

120 mujeres con diagnóstico de infarto agudo del miocardio de primera vez, incluyendo aquellas fallecidas durante la hospitalización

*Controles*

Mujeres aseguradas durante el periodo de estudio; se seleccionaron al azar siete controles por caso entre las mujeres que estuvieron aseguradas entre 1978 y 1984 y que utilizaron la farmacia durante el mismo periodo que los casos



*Exposición*

El consumo de estrógenos como terapia de remplazo hormonal en la posmenopausia fue establecido en mujeres que lo utilizaron al menos 12 meses en forma continua desde la prescripción médica. La exposición se obtuvo de la lista de usuarias de la farmacia

*Resultados*

Consumo de estrógenos conjugados en los casos: 15%

Consumo de estrógenos conjugados en los controles: 18%

Incidencia de infarto agudo del miocardio en mujeres consumidoras de estrógenos conjugados: 6.8

Incidencia de infarto agudo del miocar`dio en mujeres no consumidoras de estrógenos conjugados: 9.7

Razón de momios para infarto agudo del miocardio entre consumidoras y no consumidoras de estrógenos conjugados fue: RM 0.7; IC 95% 0.4-1.3

*Conclusiones*

Los datos sugieren que la terapia de remplazo hormonal con estrógenos conjugados se asocia en forma inversa con el riesgo de sufrir infarto agudo del miocardio (estadísticamente no significativa) y es consistente con otros estudios

Fuente: referencia 30

mitado al estudio de los factores asociados con la exposición al factor infeccioso y la determinación de cómo la exposición a un agente infeccioso ocurre es más eficiente y no sesgada que la utilización de un tradicional estudio de casos y controles donde la determinación de la exposición ocurre después de que

**Cuadro VII Ejemplo de estudio caso-autocontrol**

*Antecedente*

La intuición y diversos estudios con limitaciones metodológicas sugieren que eventos de máximo estrés y "enojo" preceden inmediatamente al desarrollo de un infarto agudo del miocardio (IAM)

*Material y métodos*

Entrevista a 1 623 sujetos, en promedio cuatro días después de un infarto agudo del miocardio

*Exposición*

1. Características del evento clínico
2. Frecuencia de enojo y estrés durante el año previo
3. Intensidad del enojo, estrés y otros factores desencadenantes 26 horas antes del IAM
4. "Enojo" fue cuantificada con una escala de estrés (autorreporte de siete niveles)
5. Consumo de aspirina

*Diseño*

Caso-caso, mediante comparación de la ocurrencia de enojo intenso dos horas previas a la ocurrencia del IAM en relación con dos autocontroles ("self-matched") pareados

*Resultados*

La razón de momios de IAM dos horas después de un episodio de enojo fue: RM = 2.3; IC 95% 1.7-3.2

Los usuarios regulares de aspirina tuvieron una probabilidad menor que los no usuarios (RM = 2.9; IC 95% = 2.0-4.5, $p < 0.05$)

*Conclusiones*

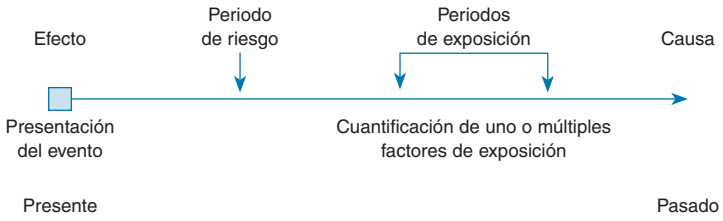
Episodios de máximo estrés y enojo son capaces de desencadenar un evento de IAM, pero la aspirina parece reducir esta probabilidad

Fuente: referencia 32

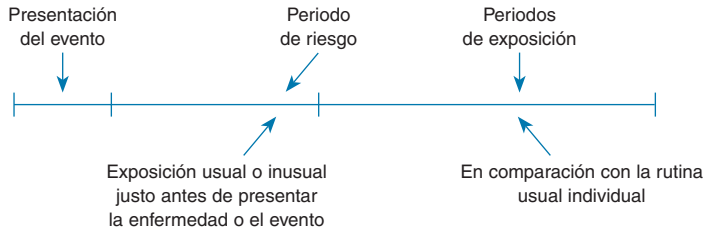
una exposición a un agente infeccioso no puede ser estudiada.

El problema de comparar casos de una enfermedad con casos de otra enfermedad proviene del hecho que las enfermedades son diferentes, incluyendo los factores de exposición.[34] Sin embargo, este problema puede desaparecer si los casos de una enfermedad específica son comparados con ellos mismos. Esto generalmente no es útil cuando no existen diferencias reales entre los grupos. En el caso de enfermedades infecciosas, sin embargo, es posible conformar diversos subgrupos de enfermedad; a pesar de que cuentan con la misma predisposición del huésped, similares métodos de detección y exposición a factores de riesgo, ellos difieren por el subtipo de organismo que es causa de infección. Un ejemplo de ello es la salmonella que afecta diferentes productos alimenticios en diversos grados de acuerdo al subtipo específico. Por ejemplo la salmonella enteritidis es común en huevos y aves de corral pero no en el ganado. Salmonella *typhimurium* es más común en cerdos. En este contexto, detallar subtipos influye en el conocimiento de patrones de resistencia a antibióticos, o en la posible identificación de cepas que pueden te-

a) Dirección de la investigación

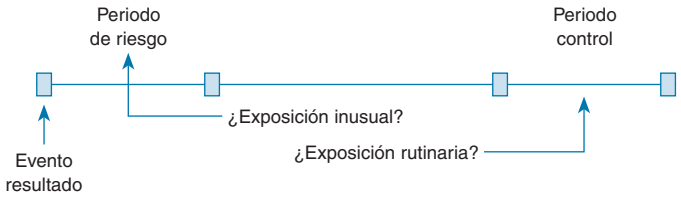



b) Comparación de exposición en el mismo



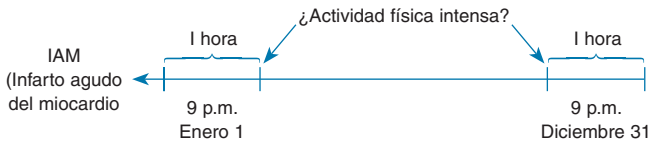



c) Periodos de exposición




d) Ejemplo: infarto agudo del miocardio y esfuerzo físico intenso

Actividad física intensa en una hora periodo en comparación con la misma hora periodo del dia previo a la presentación del evento.




El sentido del tiempo va de derecha a izquierda

**Figura 5** **Diseño caso-autocontrol**

**Cuadro VIII Ejemplo de una variante de estudio de caso-caso con un grupo control**

*Antecedentes*

La incidencia de infecciones nosocomiales causadas por patógenos gram negativos resistentes a antibióticos se ha incrementado durante los últimos años.

*Objetivo*

Evaluar simultáneamente los factores asociados a resistencia o susceptibilidad al antibiótico ceftriaxone en cultivos positivos de citrobacter freundii en un centro de atención de tercer nivel en un periodo de tres años

*Métodos*

Utilización de un estudio de caso-caso-control. Identificación de todos los sujetos positivos a citrobacter freundii admitidos entre 1997 y 2000

El primer grupo de casos fueron sujetos con cultivos clínicos nosocomiales resistentes al ceftriaxone (RC)

El grupo control fueron los sujetos que fueron admitidos al hospital en el mismo periodo de tiempo y en el mismo servicio (frequency matched controls), en una relación de seis controles por cada caso de RC y SC (susceptibles al ceftriaxone)

*Resultados*

Los principales factores asociados a RC fueron el antecedente de estancia en terapia intensiva (RM = 2.6 IC95% 1.1-6.2), exposición previa a antibioterapia (imipenem, cefalosporinas, vacomicina o tazobactam) y una estancia hospitalaria mayor de 1 semana (RM = 3.6 IC95% 1.3-10.2). Los factores asociados a SC fueron condiciones comórbidas como los diagnósticos de sida (RM = 9.5 IC95% 1.6-55.5), enfermedad vascular periférica y cerebrovascular, entre otros

*Conclusiones*

El aislamiento de cultivos de citrobacter freundii resistentes al ceftriaxone estuvo fuertemente asociado al consumo previo de antibióticos de amplio espectro y prolongada estancia hospitalaria, mientras que los cultivos con susceptibilidad al ceftriaxone no incluyeron exposición previa a antibióticos pero sí a condiciones co-mórbidas

Fuente: referencia 35



ner una fuente común. De esta manera los casos que pueden tener una exposición común pueden ser definidos y comparados con otros casos que tienen un diferente subtipo infeccioso. Este es un ejemplo de una principal utilidad, la vigilancia epidemiológica.

El efecto de comparar un grupo de casos con la misma enfermedad puede ser útil para disminuir el sesgo de selección de cuando se comparan con sujetos sanos y se cuantifica la historia de exposición actual o como dicha exposición es recordada. Paradójicamente esto puede ser una desventaja, ya que puede existir una gran variedad de exposiciones que pueden diluirse unas con otras. En el diseño caso-caso se *aisla la exposición* del agente infeccioso de otros factores etiológicos; pero otros determinantes como factores de susceptibilidad del huésped tales como edad, medicamentos usados, enfermedades crónicas y factores no conocidos, no se estudian con este tipo de diseño porque se distribuyen igualmente en los casos y el grupo "control".

Los estudios de comparación caso-caso son factibles por el avance tecnológico en las técnicas de tipificación microbiológica y un ejem-

plo de ellos es descrito en el Cuadro VIII.[35] Esta estrategia ofrece única oportunidad en términos de control de simultaneidad y el sesgo de selección que otros tipos de estudios de casos y controles pueden generar. No son útiles para determinar otros factores de riesgo en general, pero son muy rápidos y eficientes para identificar las fuentes de infecciones.

## SESGOS

El hecho de que los casos y los controles se seleccionen utilizando diferentes esquemas y que la información se obtenga en la mayoría de las veces de forma retrospectiva, hace que este tipo de estudios sean más vulnerables al efecto de diferentes sesgos. La característica que diferencia los sesgos de los errores aleatorios es que los primeros se refieren a errores que ocurren de diferente manera entre los casos y los controles, lo que hace imposible distinguir entre diferencias reales, que podrían atribuirse a la exposición, y diferencias espurias, atribuibles a errores metodológicos. Mientras que los errores aleatorios ocurren de la misma forma en ambas series y tienden a producir un sesgo hacia el valor nulo, cuando la asociación en estudio es real. En esta sección haremos una breve revisión de los sesgos que se presentan con mayor frecuencia en los estudios de casos y controles.



1. *Sesgos de selección*. Puesto que en los estudios de casos y controles se seleccionan los participantes sobre la base de la ocurrencia del evento, este tipo de estudio epidemiológico es particularmente vulnerable a los sesgos de selección; en consecuencia, se recomienda trabajar con casos incidentes y evitar que la exposición o variables asociadas a ésta determinen o condicionen la participación en el estudio, ya sea como caso o como control. Un ejemplo de este tipo de sesgos puede encontrarse en el estudio que reportó la asociación entre uso de estrógenos de remplazo y aumento en la frecuencia de cáncer endometrial. El sesgo de selección se originó en el hecho de que las mujeres que recibían estrógenos de remplazo también estaban sujetas a una vigilancia médica más estrecha, lo que producía que el diagnóstico de cáncer endometrial se realizara con mayor frecuencia en este grupo, en comparación con el grupo que no recibía estrógenos.[36] Como resultado de esta sobrevigilancia en el grupo que era usuario de estrógenos, se obtuvo una selección preferencial de los casos de cáncer endometrial que ocurrían en el grupo de mujeres expuestas a hormonales de remplazo, lo que condicionó una sobrestimación de la asociación real entre la exposición y la enfermedad.

   Los estudios de casos y controles basados en poblaciones hospitalarias, también pueden estar sujetos a sesgos de selección con relativa frecuencia. Esto ocurre si la exposición en estudio se encuentra asociada con el grupo de padecimientos seleccionado para el grupo control, siempre existirá la posibilidad de incurrir en sesgos de selección. Por ejemplo, si analizáramos la relación entre el consumo de alcohol y el riesgo de enfermedad coronaria en un estudio, en el que los casos fueran sujetos que ingresan al servicio de urgencia por enfermedad coronaria, y los controles sujetos que ingresan al mismo servicio, pero por trauma agudo, podría presentarse

un sesgo de selección si el consumo de alcohol fuera un factor que estuviera relacionado con la ocurrencia de trauma agudo. La utilización de este grupo control para estimar la frecuencia de consumo de alcohol en la población base sería incorrecta, ya que sobreestimaría el consumo y esto ocasionaría una diferencia espuria entre casos y controles.

Otra fuente de sesgo de selección puede ser la *falta de respuesta* en alguno de los grupos. Cuando esta tasa es diferente entre casos y controles y, además, difiere entre expuestos y no expuestos, se puede introducir un sesgo. En estudios donde los casos se obtienen de fuentes hospitalarias, las condiciones de invitación y convencimiento son más favorables y se logran mejores tasas de respuesta de las que generalmente se obtienen para el grupo control; sin embargo, esta tasa diferencial de participación para casos y controles sería una fuente de sesgo, siempre y cuando las tasas de participación para sujetos expuestos y no expuestos fueran diferenciales. Este tipo de sesgo puede presentarse cuando el personal de campo conoce la hipótesis en estudio y realiza un esfuerzo mayor para lograr la participación de casos o controles con la exposición de interés. Esto también puede ser una fuente de sesgo para los estudios que examinan la relación entre diferentes enfermedades y los estilos de vida. Los controles que son más conscientes de su salud y por lo tanto tienen estilos de vida más saludables, tienden a participan con mayor frecuencia en los estudios.

Otro tipo de sesgo de selección es el que se presenta al estudiar casos prevalentes, los cuales representan los sujetos con la enfermedad en estudio que sobrevivieron hasta el momento en que se inicia la investigación. En general, en este grupo se encuentran sobrerrepresentados los sujetos que cursaron con la forma más benigna de la enfermedad. Si la exposición estuviera asociada, no sólo con la ocurrencia de la enfermedad, sino también con la sobrevida, el uso de casos prevalentes podría conllevar conclusiones erróneas sobre la relación entre exposición y enfermedad.

2. *Sesgos de información.* Los estudios de casos y controles son particularmente propensos a sesgos que se introducen en el proceso de medición de la exposición, porque frecuentemente ésta se mide en forma retrospectiva, después del inicio de la enfermedad o del evento en estudio, por lo que: a) la existencia del evento puede tener un efecto directo sobre la exposición (causalidad reversa); b) la existencia del evento afecta la calidad de la medición, y c) la existencia del resultado afecta la determinación o registro de la exposición.

   Un problema importante en los estudios retrospectivos es que no se puede asegurar una relación correcta en la secuencia de la ocurrencia de la causa y el efecto. En ocasiones es posible suponer que la ocurrencia de evento puede modificar la exposición de tal forma que cuando ésta se determina no representa la medición real. Pensemos en un estudio de casos y controles hipotético, que busca comprobar la asociación entre el estado nutricional, medido éste como la concentracion de vitamina A en suero y el riesgo de cáncer de páncreas. Si se estudiaran casos avanzados de la enfermedad es muy probable que éstos tuvieran como condición de la enfermedad misma una deficiencia de vitamina A. Esto último nos podría llevar a concluir que la deficiencia de vitamina A pudiera estar

asociada con un incremento en el riesgo de cáncer de páncreas. Esta asociación espuria es producto de la enfermedad y no de una asociación real.

La información en los estudios de casos y controles frecuentemente se obtiene por medio de cuestionarios que aplican los entrevistadores, o mediante los registros médicos; por lo tanto, es común que las personas tengan problemas para recordar la información exacta sobre alguna exposición pasada. Sin embargo, es importante minimizar estas diferencias entre los casos y los controles. Por ejemplo, una mujer que tuvo un hijo con algún defecto congénito, y que es estudiada como caso, es probable que trate y haga un esfuerzo por recordar cualquier exposición durante el embarazo. En comparación, una mujer que tuvo un niño sano y que es estudiada como control, no tendrá la misma motivación para recordar, y es probable que consigne una información con un grado menor de exactitud. En este ejemplo, la respuesta es diferencial entre ambos grupos, lo que podría introducir un sesgo de información conocido como sesgo de *recordatorio.* En este ejemplo en particular, es probable que la asociación entre la exposición y el evento esté sobrestimada, debido a un mayor grado de error en la determinación de la exposición entre los controles.

Otro de los sesgos potenciales surge cuando los entrevistadores conocen la condición de caso y control, lo que puede conducir a que la entrevista se realice de manera diferencial entre los grupos. Por ejemplo, un entrevistador mal capacitado podría inducir respuestas positivas sobre la exposición preferencialmente en el grupo de casos.

Otro sesgo de información puede ocurrir al clasificar a los individuos como expuesto o no expuesto, utilizando información sobre su condición de caso o control. Esto suele ocurrir durante la medición de la exposición basada en exámenes médicos o en resultados de exámenes de laboratorio. La definición de exposición debe sujetarse a criterios estrictos y estandarizados, definidos a priori, y las decisiones o revisiones siempre deben llevarse a cabo sin el conocimiento de la condición de caso o control. Esto último podrá lograrse siempre que el laboratorio que determina la exposición se mantenga ciego o enmascarado respecto de la información sobre la condición del evento.



## ANÁLISIS E INTERPRETACIÓN

A diferencia de los estudios de cohorte, en los que puede calcularse la razón de riesgos y la razón de tasas de incidencia, en los estudios de casos y controles no puede estimarse directamente la incidencia de la enfermedad en sujetos expuestos y no expuestos. Esto se debe, primordialmente, a que los individuos se seleccionan con base en la presencia o ausencia del evento de estudio y no por el estatus de exposición (una excepción son los estudios anidados y de caso-cohorte), donde podrá estimarse la incidencia, si se conocen las fracciones muestrales de exposición en los casos y en los controles. Por esta razón, la RM (también llamada razón de ventajas, de productos cruzados, de suertes y de oportunidad relativa, entre otras) se uti-

liza como estimador para medir la asociación entre una exposición y una enfermedad.

A continuación se presentan consideraciones básicas del análisis de casos y controles, así como de sus diferentes modalidades.

**Análisis sin pareamiento**

Se basa en la tabla tradicional de 2 × 2. Es indispensable recordar que se parte de que el evento ya ocurrió y que se medirá el antecedente de exposición. Entonces, pueden calcularse los momios de exposición tanto en los casos como en los controles, esto es, comparar la posibilidad de ocurrencia de un evento con la posibilidad de que no ocurra bajo las mismas condiciones en cada uno de esos grupos, y se expresa de la siguiente manera:

Momios de exposición en el grupo de los casos: a/b
Momios de exposición en el grupo de los controles: c/d

Comparando los momios de ocurrencia de la exposición en los casos y en los controles, obtenemos la razón de momios:

$$RM = \frac{\text{momios de exposición en los casos}}{\text{momios de exposición en los controles}} = \frac{a/b}{c/d} = \frac{a \times b}{b \times c}$$

La interpretación de los resultados es la siguiente: si la RM es igual a uno, la exposición no está asociada con el evento o enfermedad; si la RM es menor de uno, la exposición está asociada de manera inversa con el evento, esto es, la exposición disminuye la posibilidad de desarrollar el evento; si la RM es mayor de uno, la exposición se encuentra asociada positivamente con el evento, lo que quiere decir que la exposición aumenta la posibilidad de desarrollar el evento.

Para cuantificar la precisión de la asociación, se realiza el cálculo de los intervalos de confianza, normalmente estimados para un nivel de confianza de 95%, como se observa en el cuadro IX; esto es, si se repitiera el mismo estudio *n* veces, bajo las mismas suposiciones estadísticas, en 95% de los casos el estimador puntual de la RM estará contenido dentro de los límites estimados.

Cuando la RM tiene valor por arriba del valor nulo (uno) y los intervalos de confianza no abarcan dicho valor, puede calcularse el impacto de la exposición mediante el riesgo atribuible (llamado también fracción atribuible); en otras palabras, la proporción de la enfermedad que se evitaría si se lograra erradicar la exposición. Consideremos como ejemplo un estudio de casos y controles realizado en la Ciudad de México para evaluar la asociación entre la obesidad y el cáncer de endometrio.[37] En dicho estudio, 84 casos confirmados histopatológicamente fueron comparados con 626 controles, seleccionados aleatoriamente de la misma fuente de obtención de los casos. Para efecto de este ejercicio, la obesidad se definió como un índice de masa corporal (IMC) mayor de 30 puntos (cuadro X).

La obesidad se presentó en 35 mujeres en el grupo de los casos y en 166 en el grupo control. El resultado es una prevalencia de obesidad de 41% y 26%, respectivamente. Al evaluar la relación entre la obesidad y el cáncer endometrial se encuentra una asociación positiva: la probabilidad de padecer cáncer endometrial fue 1.98 veces mayor en las mujeres con IMC > 30 (con obesidad), comparada con mujeres que tuvieron un IMC ≤30 (sin obesidad); esta asociación fue significativa, ya que los IC95% no abarcaron el valor nulo, oscilando la variabilidad de esta asociación de 1.24 hasta 3.16. Puesto que fue positiva y estadísticamente significativa, pudo evaluarse el impacto de

## Cuadro IX Análisis clásico de un estudio de casos y controles no pareado para evaluar razón de momios

| | *Exposición* | | |
|---|---|---|---|
| | *Sí* | *No* | *Total* |
| Casos | a | b | $n_1$ |
| Controles | c | d | $n_0$ |
| Total | $m_1$ | $m_0$ | N |

Casos: sujetos que presentan el evento (enfermedad)
Controles: sujetos que no presentan el evento

*Resultado*

| | |
|---|---|
| Prevalencia de exposición en los casos: | $a/n_1$ |
| Prevalencia de exposición en los controles: | $c/n_0$ |
| Momios de exposición en los casos: | $a/b$ |
| Momios de exposición en los controles: | $c/d$ |
| Razón de momios (RM): | $a * d/b * c$ |
| IC 95%: | $e^{\ln(RM) \pm 1.96 * EE}$ |
| Error estándar (EE): | $\sqrt{1/a + 1/b + 1/c + 1/d}$ |
| Riesgo atribuible poblacional (Rap o RAPP): | $a/n_1\ (RM - 1)/RM$ |
| Riesgo atribuible en los expuestos (Rae o RAPexp): | $RM - 1/RM$ |

*Categoría de referencia*

a: sujetos con el evento y que estuvieron expuestos
b: sujetos con el evento y que *m* estuvieron expuestos
c: sujetos que *m* presentaban el evento y que estuvieron expuestos
d: sujetos que *m* presentaban el evento y que *m* estuvieron expuestos
$m_1$: total de sujetos expuestos
$m_0$: total de sujetos no expuestos
$n_1$: total de casos
$n_0$: total de controles
N: total de la población en estudio

*ln: logaritmo natural*

**Cuadro X Cálculo de la razón de momios para evaluar el riesgo de cáncer endometrial con relación a obesidad**

| | *Exposición* | | |
|---|---|---|---|
| | *Expuestos* | *No expuestos* | |
| | *>30* | *≤30* | *Total* |
| Casos | 35 | 49 | 84 |
| Controles | 166 | 460 | 626 |
| Total | 201 | 509 | 710 |

Hipótesis

La obesidad (índice de masa corporal, IMC, >30) es un factor de riesgo para cáncer endometrial

*Diseño*

Un grupo de mujeres con edades entre 22 y 79 años con cáncer endometrial se compara con un grupo de mujeres sin la enfermedad. Todas las mujeres fueron categorizadas con obesidad de acuerdo con el IMC: obesidad, IMC > 30; sin obesidad, IMC ≤ 30

*Resultados*

| | |
|---|---|
| Prevalencia de obesidad en los casos: | $35/84 = 0.42$ |
| Prevalencia de obesidad en los controles: | $166/626 = 0.26$ |
| Momios del grupo de los casos: | $35/49 = 0.71$ |
| Momios del grupo de los controles: | $166/460 = 0.36$ |
| Razón de momios (RM): | $35 * 460/49 * 166 = 1.98$ |
| IC 95% para la razón de momios: | $e^{\ln 1.98 \pm 1.96 * 0.2391} = 1.24 - 3.16$ |
| Error estándar (EE): | $\sqrt{1/35 + 1/49 + 1/166 + 1/460} = 0.2391$ |
| Riesgo atribuible poblacional (Rap): | $35/84 * 1.98 - 1/1.98 = 0.21$ |
| Riesgo atribuible en los expuestos (Rae): | $1.98 - 1/1.98 = 0.49$ |

Fuente: referencia 37

la obesidad sobre el cáncer de endometrio. En este caso, el riesgo atribuible en la población fue de 0.21; en otras palabras, la obesidad en la población general fue responsable de 21% de los casos de cáncer de endometrio.

### Análisis con pareamiento individual

Cuando el diseño de los estudios de casos y controles contempla un pareamiento individual por algún factor de confusión, el análisis tiene particularidades diferentes al de los estudios tradicionales de casos y controles. La tabla de 2 × 2 adquiere una connotación diferente. Dada esta condición, la razón de momios pareada ($RM_p$) puede calcularse tomando en consideración las parejas con casos expuestos y controles no expuestos y dividirlos entre las parejas con casos no expuestos y controles expuestos (b/c), es decir, se utilizan únicamente las parejas discordantes

en cuanto a la exposición. Nótese que, aunque la notación en el cuadro es la misma que los estudios no pareados (a, b, c, d), el contenido de cada celda difiere debido a que se estudian parejas. Este cálculo de la $RM_p$ considera solamente pares discordantes, y se explica por el hecho de que los pares en los que caso y control estuvieron expuestos, o en los que ambos no estuvieron expuestos, no contribuyen con información acerca de la posible asociación entre la exposición y la enfermedad (cuadro XI).

Consideremos como ejemplo un estudio de casos y controles realizado en la Ciudad de México, con 84 casos de cáncer de ovario no epitelial confirmado histopatológicamente, y 84 controles sin la enfermedad, pareados individualmente por edad, seleccionados aleato-

**Cuadro XI Análisis de un estudio de casos y controles pareado individualmente para evaluar razón de momios**

| | | *Controles* | | |
|---|---|---|---|---|
| | | *Expuestos* | *No expuestos* | *Total* |
| | Expuestos | a | b | a + b |
| Casos | No expuestos | c | d | c + d |
| | Total | a + c | b + d | n = a + b + c + d |

*Casos*
Sujetos que desarrollaron el evento (enfermedad)

*Controles*
Sujetos que no desarrolaron el evento

*Resultados*
Razón de momios pareada: $RMp = b/c$
IC 95%: $e^{\ln(b/c) \pm 1.96 \sqrt{(1/b + 1/c)}}$

*Categoría de referencia*
a: parejas con caso expuesto y control expuesto
b: parejas con caso expuesto y control no expuesto
c: parejas con caso no expuesto y control expuesto
d: parejas con caso no expuesto y control no expuesto
a + c: total de parejas con controles expuestos
b + d: total de parejas con controles no expuestos
a + b: total de parejas con casos expuestos
c + d: total de parejas con casos no expuestos
n: total de parejas en el estudio

riamente de la fuente de la cual fueron obtenidos los casos.[38] La exposición estudiada fue la paridad (por lo menos un parto a término), y se consideraron no expuestas aquellas mujeres que nunca habían tenido un parto a término.

Las parejas formadas por caso y control expuesto (concordantes) sumaron 30 (celda a); y las parejas formadas por caso y control no expuestos (concordantes) sumaron 24 (celda d); éstas se excluyeron para el cálculo, y se utilizaron exclusivamente las parejas discordantes, por lo que se obtuvo una $RM_p = 0.25$. Esto significa que la paridad se encuentra asociada de manera inversa al cáncer de ovario no epitelial, interpretándose como que las mujeres que tuvieron, por lo menos, un parto a término, tienen cuatro veces menos posibilidades de padecer cáncer ovárico no epitelial, al compararlas con aquellas mujeres que nunca tuvieron un parto a término (el valor resulta de invertir la $RM_p$ solamente para facilitar la interpretación: $1/0.25 = 4$; por lo tanto, la interpretación se realiza con base en el cambio en la exposición).

Los intervalos de confianza, al no abarcar el valor nulo, hacen que la asociación sea significativa ($RM_p = 0.25$; IC95% = 0.10 – 0.61) (cuadro XII).

**Análisis de mortalidad proporcional**

Una variante de diseños de casos y controles son los estudios de mortalidad proporcional. En este caso se utiliza la información de registros o censos sobre causas de mortalidad por alguna enfermedad, por ejemplo, de cáncer; sin embargo, los mismos registros no permiten obtener información completa sobre una característica que puede ser de interés para el estudio (denominadores) o puede faltar información sobre los datos de los casos (numeradores). En tal situación, principalmente cuando no se cuenta con información acerca de los denominadores, la opción más apropiada es realizar un estudio de mortalidad proporcional, donde el cálculo de la razón de momios de la mortalidad proporcional ($RM_{mp}$) puede brindar información valiosa de registros o estudios donde no se cuenta con denominadores.[39] Ésta puede calcularse, por ejemplo, al estimar la posibilidad de morir por un tipo de cáncer (casos) en relación con la de morir por otros tipos de cáncer (controles), en un grupo expuesto de la población, comparado con un grupo no expuesto de la misma población. Tal como se mostró previamente, en este caso, la RM sería un buen estimador de la razón de tasas de incidencia.

## CONCLUSIONES

En este artículo se han descrito las diversas alternativas y variantes de los diseños de casos y controles, cuya elección depende de la hipótesis de estudio, de la exposición y tipo de evento, del tipo de información disponible y del conocimiento metodológico y analítico, así como de la imaginación e intuición del investigador para plantear un diseño de estudio y dar una respuesta válida y costo-efectiva a la pregunta de la investigación. A este respecto, es necesario destacar que, si las variables en estudio son dependientes del tiempo, los diseños de estudio anidados pueden resultar ideales. Adicionalmente, cuando se tiene la posibilidad de evaluar una exposición de riesgo inusual, previa a un evento de estudio, puede utilizarse un estudio de caso-caso; o cuando se dispone de registros poblacionales de enfermedad, podrían ser eficientes los estudios de mortalidad proporcional.

**Cuadro XII Cálculo de la razón de momios pareada para evaluar el riesgo de cáncer de ovario no epitelial con relación a paridad**

| | | *Controles* | | |
|---|---|---|---|---|
| | | *Paridad* | *Sin paridad* | *Total* |
| | Paridad | 30 | 6 | 36 |
| Casos | Sin paridad | 24 | 24 | 48 |
| | Total | 54 | 30 | 84 |

*Hipótesis*

La paridad es un factor de riesgo para el cáncer de ovario de origen no epitelial

*Diseño*

Un grupo de mujeres con cáncer de ovario no epitelial es comparado con un grupo de mujeres sin cáncer, se parearon individualmente por edad tres controles por cada caso. Todas las mujeres son categorizadas con *paridad,* cuando habían tenido por lo menos un parto a término, y *sin paridad* cuando nunca lo habían tenido

*Resultados*

Razón de momios pareada: (RMp): $6/24 = 0.25$

IC 95%: $e^{\ln 0.25 \pm 1.96\sqrt{0.208}} = 0.10 - 0.61$

*Categoría de referencia*

a: parejas con caso expuesto y control expuesto = 30
b: parejas con caso expuesto y control no expuesto = 6
c: parejas con caso no expuesto y control expuesto = 24
d: parejas con caso no expuesto y control no expuesto = 24
a + c: total de parejas con controles expuestos = 54
b + d: total de parejas con controles no expuestos = 30
a + b: total de parejas con casos expuestos = 36
c + d: total de parejas con casos no expuestos = 48
n: total de parejas en el estudio = 84

Fuente: referencia 38

Finalmente, puede decirse que los estudios de casos y controles, al igual que otros estudios observacionales, están sujetos a la acción de diferentes sesgos, por lo que no tienen como principal objetivo generalizar sus hallazgos, sino apoyar relaciones causa-efecto que tendrán que ser verificadas mediante estudios analíticos con mayor poder en la escala de causalidad.

## REFERENCIAS


1. Breslow NE, Day NE. Statistical methods in cancer research. The design and analysis of cohort studies. Vol. II. Lyon, Francia: International Agency for Research on Cancer, 1994.
2. Paneth N, Susser E, Susser M. Origins and early development of the case-control study: Part 1, Early evolution. Soz Praventiv Med 2000;47: 282-288.
3. Snow J. On the mode of communication of cholera. 2nd edition, much enlarged. London: John Churchill, 1855.
4. Paneth N. Susser E. Susser M. Origins and early development of the case-control study: Part 2, The case-control study from Lane-Claypon to 1950. Soz Praventiv Med 2002; 47(6):359-365.
5. Cornfield J. A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. J Natl Cancer Inst 1951; 11:1269-1275.
6. Mantel N, Haenszel W. Statistical aspects of the analisis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22(4):719-748.
7. Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976; 104:609-620.
8. Aznar-Ramos R, Lara-Ricalde R. Cardiovascular diseases and oral contraception. Study of cases and controls. Gac Med Mex 1984;120(3):117-127.
9. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. BMJ 1952;2:1271-1286.
10. Herbst AI, Ulfelder H, Poskancer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 1971;284:878-881.
11. Ahlbom A, Staffan N. Fundamentos de epidemiología. 4ª. ed. México: Siglo XXI, 1993.
12. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol 2004; 160: 301-305.
13. Kyrklund-Blomberg NB, Granath F, Cnattingius S. Maternal smoking and causes of very preterm birth. Acta Obstet Gynecol Scand 2005 Jun;84(6):572-577.
14. Muñoz N, Bosch FX, de Sanjose S, Tafur L, Izrzugaza I, Gili M *et al.* The causal link between human papillomavirus and invasive cervical cancer: A population-based case-control study in Colombia and Spain. Int J Cancer 1992;52:743-749.
15. Stopkova P, Vevera J, Paclt I, Zukov I, Papolos DF, Saito T, Lachman HM. Screening of PIP-5K2A promoter region for mutations in bipolar disorder and schizophrenia. Psychiatr Genet 2005 Sep;15(3):223-227.
16. Romieu I, Hernández M, Lazcano E, López L, Romero R. Breast cancer and lactation history in Mexican women. Am J Epidemiol 1996;143: 543-552.
17. Pérez-Padilla R, Regalado J, Vedal S, Pare P, Chapela R, Sansores R, *et al.* Exposure to biomass smoke and chronic airway disease in Mexican women. A case-control study. Am J Respir Care Med 1996;154:701-706.
18. Hernández M, Lazcano E, Alonso P, Romieu I. Evaluation of the cervical cancer screening programme in Mexico: A population based case control study. Int J Epidemiol 1998;27:370-376.
19. Carrique Mas J, Andersson Y, Hjertqvist M, Svensson A, Torner A, Giesecke J. Risk factors for domestic sporadic campylobacteriosis among young children in Sweden. Scan J Infect Dis 2005;37(2):101-110.
20. Gelatti U, Covolo L, Franceschini M, Pirali F, Tagger A, Ribero ML, *et al.* Coffee consumption reduces the risk of hepatocellular carcinoma independently of its aetiology: a case-control study. J Hepatol 2005; 42(4):528-534.
21. Ward EM, Kramer S, Meadows AT. The efficacy of random digit dialing in selecting matched controls for a case-control study of pediatric cancer. Am J Epidemiol 1984;120(4):582-591.
22. Chiaffarino F, Pelucchi C, Negri E, Parazzini F, Franceschi S, Talamini R, *et al.* Breastfeeding and the risk of epithelial ovarian cancer in an Italian population. Gynecologic Oncology 2005;98:304-308.

23. Cullen M, Nolan J, Cullen M, Molones M, Kearney J, Lambe J, *et al.* Effect of high levels of intense sweetener intake in insulin dependent diabetics on the radio of dietary sugar to fat: a case-control study. European Journal of Clinical Nutrition 2004;58:1336-1341.
24. Cerhan JR, Cantor KP, Williamson K, Lynch CF, Torner JC, Burmeister LF. Cancer mortality among Iowa farmers: recent results, time trends, and lifestyle factors (United States). Cancer Causes Control 1998;9(3):311-319.
25. Gordis L. Epidemiology. Philadelphia: W.B. Sanders Company, 1996.
26. Calin GA, Trapazo F, Shimizu M, Dumitru CD, Yendamuri S, Godwin A, *et al.* Familial cancer associated with a polymorphism in ARLTS1. N Engl J Med 2005;352:1667-1676.
27. Dong JYS, Ho TP, Kan CK. A case control study of 92 cases of in patient suicides. Journal of Affective Disorders 2005;87:91-99.
28. Schuurman AG, Goldbohm RA, Dorant E, van den Brandt P. Anthropometry in relation to prostate cancer risk in the Netherlands Cohort Study. Am J Epidemiol 2000;151(6):541-549.
29. Nomura A, Stemmermann GN, Chyou PH, Kato I, Perez-Perez GI, Blaser MJ. *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med 1991;325(16):1132-1136.
30. Hernández AM, Walker AM, Jick H. Use of replacement and the risk of myocardial infarction. Epidemiology 1990;1:128-133.
31. Maclure M. The case-crossover design: A method for studying transient effects on the risk of acute events. Am J Epidemiol 1991;133:144-153.
32. Mittleman MA, Maclure M, Sherwood JB, Mulry RP, Tofler GH, Jacobs SC *et al.* Triggering of acute myocardial infarction onset by episodes of anger. Determinants of myocardial infarction onset study investigators. Circulation 1995;92(7):1720-1725.
33. Rothman KJ, Greenland S. Modern epidemiology. 2a. ed. Filadelfia: Lippincott-Raven Publishers, 1998:93-114.
34. McCarthy N, Giesecke J. Case case comparisons to study causation of common infectious diseases. Int J Epidemiol 1999;28:764-768.
35. Kim PW, Harris AD, Roghmann MC, Morris JG, Strinivasan A, Perencevich. Epidemiological risk factors for isolation of ceftriaxone-resistent versus susceptible citrobacter freundii in hospitalized patients. Antimicrob Agents Chemother 2003;47(9):2882-2887.
36. Horwitz RI, Feinstein AR. Alternative analytic methods for case-controls studies of estrogens and endometrial cancer. N Engl J Med 1978;299:1089-1094.
37. Salazar E, Lazcano E, Gonzalez G, Escudero P, Salmeron J, Hernández M. Case-control study of diabetes, obesity, physical activity and risk of endometrial cancer among Mexican women. Cancer Causes Control 2000;11:707-711.
38. Sánchez LM, Salazar E, Lazcano E, González G, Escudero P, Hernández M. Factors associated with non-epithelial ovarian cancer among Mexican women: A matched case-control study. Int J Gynecol Cancer 2003;13(6):756-763.
39. Miettinen OS, Wang JD. An alternative to the proportionate mortality ratio. Am J Epidemiol 1981;114:144-148.

# Estudios de casos y controles — Ejercicios

Las preguntas de este ejercicio se basan en el siguiente artículo:

## REFERENCIA

Lazcano-Ponce EC, Hernández-Ávila M, López-Carrillo L, Alonso de Ruiz P, Torres-Lobatón A, González-Lira G, Romieu I. Factores de riesgo reproductivo e historia de vida sexual asociados a cáncer cervical en México. Rev Invest Clin 1995;4: 377-385.

## RESUMEN


### Introducción

El cáncer cervicouterino (CaCu) es un problema de salud pública en México. La mortalidad por este tumor ha permanecido estable en los últimos 15 años y, de acuerdo con el Registro Histopatológico de Neoplasias Malignas para el año de 1993, este cáncer correspondía a 23.5% de todos los tumores reportados. Son varios los factores de riesgo que se han asociado con este cáncer. En el estudio, los autores evaluaron los principales factores de riesgo reproductivo de CaCu en una zona de alta incidencia.

### Material y métodos

Los autores realizaron un estudio de casos y controles en el área metropolitana de la Ciudad de México desde agosto de 1990 hasta diciembre de 1992.

### Obtención de los casos

Se entrevistaron 745 mujeres con sospecha clínica y citológica de CaCu que fueron referidas a la preconsulta de ocho hospitales (cuatro de seguridad social, dos privados y dos que brindaban atención a población abierta). Los criterios de inclusión correspondieron a: i) casos nuevos de reciente diagnóstico histopatológico, ii) que no hubieran recibido tratamiento previo, iii) habitantes de la Ciudad de México con un año, por lo menos, de residencia, y iv) que hubieran firmado la carta de consentimiento. La muestra se conformó de 630 casos, de los cuales 397 correspondieron a cáncer cervical invasor y 233 a neoplasia cervical intraepitelial grado III.

### Obtención de controles

Se llevó a cabo una selección aleatoria de 1 005 controles, por medio del marco muestral de viviendas de la Dirección General de Epidemiología de la SSA. Se utilizaron los mismos criterios de inclusión de los casos. La selección se realizó por grupos quinquenales tomando como referencia la distribución de edad de los casos de CaCu reportados al Registro Nacional de Cáncer durante el año de 1989. La tasa de respuesta de los controles fue de 84.3 por ciento.

## RESULTADOS

**Cuadro I** **Modelos multivariados de factores reproductivos e historia de vida sexual en relación con cáncer cervical en la Ciudad de México**

| | *Casos* | *%* | *Controles* | *%* | *RM** | *IC95%* | | | *Prueba de tendencia* |
|---|---|---|---|---|---|---|---|---|---|
| Edad de inicio de vida sexual | | | | | | | | | |
| ≥25 | 31 | 4.9 | 131 | 13.0 | 0.41 | 0.25 | – | 0.69 | <0.001 |
| 20-24 | 146 | 23.2 | 303 | 30.1 | 0.77 | 0.54 | – | 1.08 | |
| 19 | 48 | 7.6 | 94 | 9.4 | 0.72 | 0.46 | – | 1.14 | |
| 18 | 97 | 15.4 | 139 | 13.8 | 1.00 | | | | |
| 17 | 78 | 12.4 | 105 | 10.4 | 1.01 | 0.67 | – | 1.52 | |
| 16 | 70 | 11.1 | 85 | 8.5 | 1.01 | 0.66 | – | 1.55 | |
| 15 | 91 | 14.4 | 86 | 8.5 | 1.20 | 0.79 | – | 1.81 | |
| ≤14 | 67 | 10.6 | 59 | 5.9 | 1.23 | 0.78 | – | 1.95 | |
| Sin datos | 2 | 0.3 | 3 | 0.3 | | | | | |

* Razón de momios ajustada por edad, nivel socioeconómico, número de partos vaginales y número de parejas sexuales.



**Cuadro II** **Nivel de tabaquismo en relación con cáncer cervical en México**

| | *Casos* | *%* | *Controles* | *%* | *RM** | *IC95%* | | | *Prueba de tendencia* |
|---|---|---|---|---|---|---|---|---|---|
| Índice de tabaquismo | | | | | | | | | |
| No fumador | 458 | 72.7 | 810 | 80.6 | 1.00 | | | | |
| 1er. tercil | 70 | 11.1 | 69 | 6.9 | 1.42 | 0.96 | – | 2.09 | |
| 2o. tercil | 53 | 8.4 | 67 | 6.7 | 1.18 | 0.78 | – | 1.79 | |
| 3er. tercil | 49 | 7.7 | 59 | 5.9 | 1.25 | 0.81 | – | 1.93 | >0.05 |

* Razón de momios ajustada por edad, nivel socioeconómico, edad de inicio de vida sexual, número de partos vaginales y número de parejas sexuales.

**Cuadro III** **Evaluación del programa de tamizaje de cáncer cervical en México. Historia de realización del examen de Papanicolau (Pap) y casos de cáncer cervical in situ**

| | *Casos* | *Controles* | *OR*** |
|---|---|---|---|
| Sin Pap previo | 87 | 485 | 1.00 |
| Pap con síntomas ginecológicos | 135 | 499 | 1.55 |
| Pap sin síntomas ginecológicos y sin resultados del Pap | 49 | 412 | 0.68 |
| Pap sin síntomas ginecológicos con resultados del Pap | 41 | 353 | 0.66 |

**Cuadro IV** **Evaluación del programa de tamizaje de cáncer cervical en México. Cáncer cervical invasor**

| | *Casos* | *Controles* | *OR*** |
|---|---|---|---|
| Sin Pap previo | 224 | 485 | 1.00 |
| Pap con síntomas ginecológicos | 165 | 499 | 0.76 |
| Pap sin síntomas ginecológicos y sin resultados del Pap | 69 | 412 | 0.38 |
| PAP sin síntomas ginecológicos con resultados del Pap | 43 | 353 | 0.28 |

** Razones de momios ajustadas por edad al inicio de la vida sexual, número de nacimientos normales, número de parejas sexuales y nivel socioeconómico.



## PREGUNTAS

1. ¿Qué tipo de estudio epidemiológico recomendaría para alcanzar el objetivo que plantean los autores? Explique por qué.
2. Mencione el evento de interés y su escala de medición.
3. ¿La identificación y selección de los casos y controles fue influida por el estado de exposición?
4. ¿Considera usted que el grupo control fue adecuado para estimar los factores de exposición en la población base que dio origen a los casos?

Para contestar las preguntas 5 a 8 obtenga los datos del cuadro I, que contiene información extraída del artículo. Para facilitar sus estimaciones, llene la tabla de 2 × 2 que a continuación se le presenta. Para esto, no tome en cuenta a las mujeres sin datos.

| | Edad en años inicio de vida sexual | | |
|---|---|---|---|
| | <20 | ≥20 | Total |
| Casos | | | |
| Controles | | | |
| Total | | | |

5. ¿Qué proporción de mujeres con edad de inicio de vida sexual menor de 20 años se observó entre los casos y qué proporción entre los controles? (No tome en cuenta a las mujeres sin datos). Interprete.

De acuerdo con los datos del presente trabajo, las mujeres presentan un mayor riesgo de CaCu conforme más temprano inician su vida sexual. Calcule lo que a continuación se le solicita:

6. Estime los momios de enfermar en el grupo expuesto (mujeres que iniciaron su vida sexual activa antes de los 20 años de edad). Interprete.
7. Estime los momios de enfermar en el grupo no expuesto (mujeres que iniciaron su vida sexual activa después de los 19 años (≥ 20 años)). Interprete.
8. Estime la razón de momios. Interprete.



Para contestar las preguntas 9 a 12 obtenga los datos del cuadro II, que contiene información extraída del artículo. Para facilitar las estimaciones llene la siguiente tabla:

| | Tabaquismo | | |
|---|---|---|---|
| | Sí | No | Total |
| Casos | | | |
| Controles | | | |
| Total | | | |

9. De acuerdo con los datos del cuadro II, calcule las prevalencias de fumadoras observadas entre los casos y entre los controles.
10. Estime los momios de enfermar en el grupo expuesto (mujeres fumadoras). Interprete.
11. Estime los momios de enfermar en el grupo no expuesto (mujeres no fumadoras). Interprete.
12. Estime la RM. Interprete.
13. De acuerdo con los datos del cuadro II, calcule el riesgo atribuible proporcional en los casos expuestos al tabaco. Aclare los pasos que empleó para realizar sus cálculos e interprete sus resultados.
14. Calcule el riesgo atribuible proporcional del tabaquismo en la población general. Aclare los pasos que empleó para realizar sus cálculos e interprete sus resultados.

Como parte del mismo trabajo se evaluó la utilidad del examen de Pap y los principales resultados se muestran en los cuadros III y IV. Responda las siguientes preguntas a partir de estos resultados.

15. Interprete el valor puntual de la RM de la cuarta categoría, donde se estima la utilidad del examen de Pap cuando las mujeres acuden sin síntomas ginecológicos y reciben resultados de citología. Para cáncer in situ (RM = 0.66) e invasor (RM = 0.28).
16. Estime la proporción prevenible de casos de cáncer cervical in situ e invasor, atribuibles a la realización del examen de Pap dentro de la cuarta categoría (acuden sin síntomas y reciben resultados).

## RESPUESTAS

1. Los autores realizaron un estudio de casos y controles, debido a que se trata de una patología rara y con largo periodo de latencia.
2. El evento de interés corresponde a cáncer cérvico uterino, escala nominal (dicotómica).
3. No, la selección de los casos y controles fue independiente de la exposición. Fueron varios los factores de riesgo de cáncer cérvicouterino que se tomaron en cuenta, de tal manera que difícilmente la selección de las mujeres estuvo determinada por alguno de los factores de riesgo.
4. Sí, ya que los controles fueron seleccionados aleatoriamente de la población de donde en teoría surgieron los casos. Es muy probable que, si las mujeres seleccionadas como controles hubieran enfermado de cáncer cérvico uterino, habrían acudido a alguno de los hospitales de donde se obtuvieron los casos.



A continuación se presenta la tabla que se construye a partir del cuadro I. Recuerde que para llenar la tabla, no debió tomar en cuenta a las mujeres sin datos.

| | Edad en años inicio de vida sexual | | |
|---|---|---|---|
| | <20 | ≥20 | Total |
| Casos | 451 | 177 | 628 |
| Controles | 568 | 434 | 1002 |
| Total | 1019 | 611 | 1630 |

5. El 71.8% (451/628) de los casos y 56.7% (568/1002) de los controles iniciaron su vida sexual activa antes de los 20 años.
6. Los momios de enfermar en el grupo expuesto (mujeres que iniciaron su vida sexual activa antes de los 20 años de edad) = (451/568) = 0.794.

*Interpretación: entre las mujeres que iniciaron su vida sexual activa antes de los 20 años de edad por cada 0.79 casos de CaCu hay un control (0.79:1).*

7. Los momios de enfermar en el grupo no expuesto (mujeres que iniciaron su vida sexual activa después de los 19 años (⩾ 20 años)) = (177/434) = 0.408.

*Interpretación: Entre las mujeres que iniciaron su vida sexual activa después de los 19 años de edad por cada 0.41 casos de CaCu hay un control (0.41:1).*

8. La razón de momios = 1.95.

*Interpretaciones:*

*a. Las mujeres de este estudio que iniciaron su vida sexual antes de los 20 años de edad, tuvieron 1.95 veces el riesgo de padecer CaCu que las que iniciaron su vida sexual después de los 20 años.*

*b. Las mujeres de este estudio que iniciaron su vida sexual antes de los 20 años, tuvieron 95 % más riesgo de padecer CaCu que las que iniciaron su vida sexual después de los 20 años.*

| | Tabaquismo | | |
|---|---|---|---|
| | Sí | No | Total |
| Casos | 172 | 458 | 630 |
| Controles | 195 | 810 | 1005 |
| Total | 367 | 1268 | |

9. El 27.3% de los casos (172/630) y 19.4% (195/1005) de los controles fumaban
10. Los momios de enfermar en el grupo expuesto (mujeres fumadoras) = (172/195) = 0.882.

*Interpretación: entre las mujeres fumadoras, por cada 0.88 casos de CaCu hay un control (0.88:1).*

11. Los momios de enfermar en el grupo no expuesto (mujeres no fumadoras) = (458/810) = 0.565.

*Interpretación: entre las mujeres no fumadoras, por cada 0.57 casos de CaCu hay un control (0.57:1).*

12. La razón de momios = 1.56.

*Interpretaciones:*

*a. Las mujeres de este estudio que estuvieron expuestas al tabaquismo tuvieron 1.56 veces el riesgo de padecer CaCu que las que no estuvieron expuestas.*

*b. Las mujeres de este estudio que estuvieron expuestas al tabaquismo tuvieron 56% más riesgo de padecer CaCu que las que no estuvieron expuestas.*

13. Para calcular el riesgo atribuible proporcional entre los expuestos (o fumadores), se utilizó información del segundo cuadro construido por usted.

Momios expuestos = 172/195 = 0.88
Momios no expuestos = 458/810 = 0.57

RM = 0.88/0.57 = 1.56
Rae = (1-(1/RR)) × 100 = 35.9

*Interpretación: 39.5% del total del riesgo de enfermar por cáncer cervical en los expuestos es atribuible al tabaquismo.*

14. Riesgo atribuible proporcional en la población general (RAPP) a fumar.

Procedimiento:

$$\text{Rap} = \frac{p_e \times (RM - 1)}{p_e \times (RM - 1) + 1} \times 100$$

donde,
$p_e$ = prevalencia de exposición



*Interpretación: La prevalencia de exposición en el ámbito poblacional se obtiene de la prevalencia de exposición en los controles, debido a que los controles representan a la población general.*

*195/1005 = 0.194*

$$\text{Rap} = \frac{0.194 \times (1.56 - 1)}{0.194 \times (1.56 - 1) + 1} \times 100 = 9.8\%$$

*Interpretación: si asumimos una prevalencia de tabaquismo poblacional de 19.4% en mujeres ≥25 años, la proporción de riesgo de enfermar de cáncer cervical asociado con el tabaco en la población total de mujeres ≥25 años es de 9.8%.*

15. Interpretación del valor puntual de la razón de momios de la cuarta categoría, donde se estima la utilidad de la prueba de Pap cuando las mujeres acuden sin síntomas ginecológicos y reciben resultados de citología. Para cáncer in situ (RM = 0.66) e invasor (RM = 0.28).

Procedimiento: 1 − 0.66 = 0.34

*Interpretación: el riesgo de padecer cáncer cervical in situ en las mujeres que acuden sin síntomas ginecológicos y reciben resultados de citología fue 34% menor que en aquellas mujeres sin Pap previo.*

Procedimiento: $1 - 0.28 = 0.72$

*Interpretación: el riesgo de padecer cáncer cervical invasor en las mujeres que acuden sin síntomas ginecológicos y reciben resultados de citología fue 72% menor que en aquellas mujeres sin Pap previo.*

16. Proporción prevenible de casos de cáncer cervical in situ e invasor atribuibles a la prueba de Pap dentro de la cuarta categoría (acuden sin síntomas y reciben resultados).

Cáncer cervical in situ

Procedimiento: fracción prevenible = (1-RM) × 100 = 34%

RM = 0.66
Inverso RM = 1.5

*Interpretación: la asistencia a la prueba de Pap sin síntomas y con recepción de resultados reduce 34% de casos de cáncer cervical in situ en comparación con no contar con un Pap previo.*



Cáncer cervical invasor

Procedimiento: fracción prevenible en los expuestos = (1-RM) × 100 = 72%

RM = 0.28
Inverso RM = 3.6

*Interpretación: la asistencia a la prueba de Pap sin síntomas y con recepción de resultados reduce 72% de casos de cáncer cervical invasor en comparación con no contar con un Pap previo.*

# VIII Encuestas transversales

*Bernardo Hernández Prado*
*Héctor Eduardo Velasco Mondragón*

La encuesta transversal es un diseño de investigación epidemiológica de uso frecuente. Se trata de estudios típicamente observacionales y son también llamados encuestas de prevalencia o estudios transversales.[1-6] El diseño de una encuesta transversal debe considerar aspectos relacionados con la población que se estudiará, los sujetos de quienes se obtendrá información y la información que se busca captar.

Las encuestas transversales (cuando se realizan como estudios descriptivos) se dirigen primordialmente al estudio de la frecuencia y distribución de eventos de salud y enfermedad, aunque también se pueden utilizar para explorar y generar hipótesis de investigación (en el caso de los estudios analíticos). En el primer caso, las encuestas tienen como fin medir una o más características o enfermedades en un punto específico en el tiempo, por ejemplo: el número de enfermos con diabetes en la población mexicana en el año 2000; el número promedio de miembros de las familias rurales que enfermaron en un periodo determinado; el promedio de edad de hombres y mujeres que utilizaron o no servicios de salud durante el último trimestre del año; el nivel de satisfacción de pacientes atendidos por médicos familiares el mes previo, o la intención en hombres y mujeres de cesar de fumar en los meses siguientes. Las encuestas transversales son de gran utilidad en lo que respecta a la gestión de la salud pública, ya que producen información muy valiosa sobre el estado de salud de la población y permiten generar hipótesis de investigación sobre su origen. Son muy útiles para estimar la prevalencia de algunos padecimientos (esto es, la proporción de individuos que padece alguna enfermedad en una población en un momento establecido), así como la distribución y frecuencia de posibles factores de riesgo para algunas enfermedades.

Cuando el objetivo es explorar hipótesis de investigación, la característica distintiva de este tipo de estudios es que la variable de resultado (enfermedad o condición de salud) y las variables de exposición (características de los sujetos) se determinan de manera simultánea en la unidad de estudio. Ejemplos de esto serían el análisis de la relación entre alcoholismo (exposición) y violencia intrafamiliar (evento); la relación entre actividad física (exposición) y obesidad (evento), y la relación entre el género (exposición) y la presencia de diabetes (evento). Es importante mencionar que, dada su naturaleza, es difícil estudiar mortalidad en estudios transversales.

A diferencia de otros diseños epidemiológicos —como los estudios de cohorte, en los cuales se realiza un seguimiento de sujetos expuestos y no expuestos para documentar la ocurrencia de eventos nuevos— en las encuestas transversales los sujetos investigados son entrevistados en un punto preciso en el tiempo y en dicho momento se obtiene simultáneamente información sobre la exposición, el evento y otras variables importantes. Debido a esta simultaneidad, no es posible determinar si el supuesto factor de exposición precedió al aparente efecto o si es producto del

mismo efecto. Esta limitación, además de las asociadas a todos los estudios observacionales, impide establecer causalidad entre exposición y efecto, salvo en el caso de exposiciones que no cambian con el tiempo y de las que, por lo tanto, sabemos que preceden el evento, por ejemplo el género, el grupo sanguíneo o el lugar de nacimiento. Esta limitación referida en algunos textos como *causalidad reversa* es una debilidad de este tipo de estudios. Las grandes limitaciones de los estudios transversales para establecer causalidad se compensan, sin embargo, por su flexibilidad, bajo costo y utilidad para la generación de datos de alto valor administrativo en la gestión y planeación de salud. No obstante, las asociaciones derivadas de este tipo de estudios deberán ser siempre objeto de verificación e interpretarse con reservas.

Las encuestas transversales se utilizan tanto para estudiar enfermedades agudas de alta frecuencia como para estudiar enfermedades de larga duración o cuyas manifestaciones se desarrollan lentamente, como es el caso de enfermedades crónicas, problemas de desnutrición o mala nutrición por exceso, etcétera. Este tipo encuestas son adecuadas para el estudio de enfermedades (o exposiciones) que se presentan con poca frecuencia en una población (enfermedades raras o de baja prevalencia) o que son de corta duración, debido a que a pesar de requerir muestras grandes arrojan datos administrativos importantes. Buenos ejemplos de este tipo de encuestas son las encuestas nacionales de salud, en las que se determina las prevalencias nacionales de enfermedades poco frecuentes a nivel poblacional pero que son de gran importancia para la vigilancia epidemiológica, como el SIDA o la hepatitis C. También son útiles para detectar enfermedades de corta duración pero de alta frecuencia, como las diarreas o las enfermedades respiratorias agudas. A pesar de sus limitaciones, los estudios transversales son comunes y útiles, ya que su costo es relativamente inferior al de otros diseños epidemiológicos y proporcionan información importante para la planificación y administración de los servicios de salud, que difícilmente pueden ser generados con estudios de cohorte o de casos y controles. Estos últimos tipos de estudios están dirigidos hacia la constatación de hipótesis, mientras que las encuestas se orientan principalmente hacia la generación de información de utilidad administrativa, como ya se mencionó anteriormente.

En el presente artículo se discuten algunos aspectos relevantes para el diseño, conducción, análisis, interpretación y aplicaciones de las encuestas transversales, así como sus ventajas y limitaciones.



## POBLACIÓN Y MUESTRA EN ESTUDIOS TRANSVERSALES

En las encuestas transversales se obtiene información sobre una población definida para fines del estudio. Se define como población base del estudio aquella a la que se extrapolarán los resultados.[7] En muchas ocasiones una encuesta transversal no obtiene información de todos los sujetos que integran la población bajo estudio, sino sobre un subgrupo de ellos llamado muestra. Al realizar la encuesta es necesario definir la unidad de observación del estudio, esto es, la unidad básica sobre la cual se captará información detallada sobre el evento de estudio, por ejemplo, individuos, familias, hogares o escuelas.

El proceso de selección de informantes es muy importante. La muestra seleccionada debe reflejar las características de la población base que se busca estudiar. Por ejemplo, para obtener la media de edad o la distribución por edades de una población, sujetos de todos los

grupos de edad deberán tener la misma oportunidad de ser incluidos en el estudio. En ocasiones el investigador puede estar interesado en estudiar características de algún subgrupo específico de su población y, por lo mismo, puede aumentar la proporción de sujetos en la muestra que pertenecen a ese subgrupo. Existen diversos métodos de selección de sujetos para participar en el estudio llamados *métodos de muestreo.* En general, los métodos de muestreo se clasifican en muestreos probabilísticos y no probabilísticos. Una muestra probabilística es aquella en la que es posible conocer la probabilidad de selección que tiene cada elemento de la población; en este caso, si bien esta probabilidad no necesariamente debe ser igual para todos los miembros, sí debe ser diferente de cero. Por el contrario, en las muestras no probabilísticas no es posible conocer la probabilidad de selección de cada elemento de la población y por consiguiente sólo son representativas de ellas mismas. Algunos procedimientos de muestreo probabilístico son el muestreo aleatorio simple, el estratificado y por conglomerados.

Los muestreos estratificado y por conglomerados tienen importantes implicaciones en el caso de las encuestas transversales. A diferencia del muestreo aleatorio simple, en el que cada elemento de la población tiene la misma probabilidad de ser seleccionado para la muestra, esta probabilidad puede variar en los procedimientos de muestreo: a) estratificado, en el cual la población se divide en estratos homogéneos antes de hacer la selección de la muestra y por lo tanto se asegura la selección de unidades en todos los estratos definidos, y b) por conglomerado, en el cual la población se subdivide en conglomerados o *clusters* (por su nombre en inglés), generalmente compuesto por unidades heterogéneas, del interior de los cuales se seleccionan las unidades de muestreo. Por ejemplo, en un muestreo por conglomerados es posible seleccionar, en la primera fase, localidades (unidades primarias de muestreo), y en la segunda, viviendas en el interior de cada conglomerado seleccionado (unidades secundarias de muestreo). Si, además, dentro de cada vivienda se selecciona un grupo de sujetos, por ejemplo para determinar la prevalencia de diabetes mediante una prueba de sangre, éstos serían las unidades terciarias de muestreo.

El muestreo por conglomerados puede introducir variación en los errores estándar de los estimadores de asociación derivados de un estudio transversal, ya que la asociación entre las unidades secundarias de muestreo puede ser distinta en el interior de cada conglomerado que entre los conglomerados. La diferencia entre los estimadores de asociación calculados que consideran el agrupamiento de observaciones y los que no lo consideran es llamada efecto de diseño.[8] Es importante considerar los efectos de diseño tanto en las estimaciones de poder y tamaño de la muestra, como en el análisis de encuestas transversales. Actualmente, algunos paquetes estadísticos como STATA o SAS cuentan con rutinas específicas para el análisis de encuestas complejas, y posibilitan el ajuste por conglomerados y estratificación.

Otro aspecto importante que debe tomarse en cuenta en el diseño de muestras para estudios transversales es el grado en el que la muestra refleja las características de la población desde su selección original. En aquellos casos en los que la muestra tiene una distribución distinta a la de la población en sus características demográficas básicas (edad, sexo, etc.), es posible aproximar la distribución poblacional mediante el uso de ponderadores, mismos que son estimados utilizando el esquema de muestreo escogido. Un ponderador es el número de observaciones en la población que representa cada observación en la muestra y está íntimamente relacionado con las probabilidades de selección de cada unidad mues-

tral. Este número, generalmente calculado en proporción inversa a la probabilidad de selección de cada elemento de la muestra, permite "expandir" la muestra para generar estimaciones en números absolutos relativos a la población de referencia. La principal consecuencia de no usar ponderadores es el cálculo impreciso del estimador poblacional. La consecuencia de no considerar el efecto de diseño introducido por el muestreo polietápico o por conglomerados consiste en la estimación equívoca de la varianza, lo que en general se traduce en subestimaciones tanto en las pruebas de hipótesis como en la determinación de los intervalos de confianza (IC).

Por ejemplo, en la Encuesta Nacional de Salud 2000 (ENSA 2000),[9] los porcentajes de diabetes mellitus sin ponderar ni considerar el tipo de muestreo fueron de 8.7 a escala global, 8.6% para los hombres y 8.8% para las mujeres. Después de la ponderación, estos valores se estimaron en 7.5 a escala global, 7.2% para los hombres y 7.8% para las mujeres. El análisis de la ENSA 2000 también muestra que la prevalencia de diabetes mellitus se encuentra asociada con el índice de masa corporal (IMC). No obstante, los valores de esta asociación se modifican según se utilicen o no ponderadores. De esta manera, si se comparan los casos con IMC igual o mayor a 30 con aquéllos en los que el índice es de 29 o menos, se obtienen razones de momios de prevalencia (RMP) ajustadas por edad y sexo, de:

RMP = 2.28 con $p < 0.01$; IC95% 1.86, 2.79 sin ponderar (en este caso el intervalo es de 0.93)

RMP = 2.52 con $p < 0.01$; IC95% 2.23, 2.86 con ponderadores (en este caso el intervalo es de 0.63)

Como puede notarse, las razones de momios de prevalencia son diferentes, y aunque los valores de $p$ son significativos, los intervalos de confianza sin ponderación resultan más amplios que los intervalos que se obtienen usando ponderadores.

Las encuestas transversales suelen sobrerrepresentar a los casos con larga duración de la enfermedad (generalmente crónica o menos severa) y a subrepresentar aquéllos de corta duración (generalmente aguda o más severa). A este sesgo de selección se le denomina sesgo de duración. Por ejemplo, en el caso de una enfermedad de duración muy variable, como el asma, una persona que contrae una enfermedad que dura desde los 20 hasta los 70 años con manifestaciones leves de la enfermedad tiene una gran posibilidad de ser incluida durante los 50 años de duración de la enfermedad; en cambio, una persona que contrae la misma enfermedad pero con manifestaciones más graves a los 20 años y que, como consecuencia, muere a la edad de 25 años, difícilmente será incluida en la muestra de estudio. Este sesgo es especialmente importante cuando la duración se asocia con la exposición en estudio.

En cuanto a la exposición, debe considerarse si ésta cambia con el tiempo (por ejemplo, consumo de tabaco, dieta, actividad física) o si no cambia (grupo sanguíneo, sexo, etc.). Otro aspecto importante a considerar, en este mismo sentido, es la ventana de causalidad relevante. Un ejemplo se presenta al investigar la asociación entre tabaquismo y algunas enfermedades específicas: el infarto, por ejemplo, se relaciona con la cantidad de cigarrillos que el paciente fumó en los últimos dos años, mientras que el cáncer se relaciona con lo que fumó en los últimos diez. Además, como mencionamos anteriormente, si la exposición se relaciona con la presencia de enfermedades cuyo curso es leve y de larga duración —aun cuando no esté asociada causalmente con el riesgo de enfermar—, la prevalencia de exposición será elevada en los casos; como consecuencia, podría

concluirse de manera errónea que existe una asociación entre la exposición y la enfermedad. Si, por el contrario, la exposición produce una enfermedad de alta letalidad, entonces es posible que la frecuencia de exposición sea baja entre los casos detectados mediante una encuesta, por lo que la asociación exposición-enfermedad podría paradójicamente ser estimada como negativa. En la primera situación descrita la exposición en los casos está sobrerrepresentada, mientras que en la segunda está subrepresentada. Los sesgos, en resumen, son errores que se concentran sistemáticamente en algún subconjunto de los sujetos de investigación y que afectan la validez interna de un estudio; es decir, disminuyen la probabilidad de que los resultados obtenidos en la muestra investigada sean correctos.

Los errores no sistemáticos afectan la validez externa e interna de los estudios. Esto significa que es más difícil que sus resultados reflejen adecuadamente lo que realmente sucede en la población que la muestra quiere representar. En general se acepta que si el modo de selección de los sujetos de estudio está relacionado con una menor o mayor exposición, o una menor o mayor proporción (o severidad) de la enfermedad en comparación con la población base, estamos cometiendo un error de selección, pues la muestra no representará adecuadamente a la población blanco y el estudio estará comprometido en lo que se refiere a su validez externa. Como hemos mencionado, los estudios transversales tienen un alto valor para la administración de los programas y servicios de salud, pero siempre y cuando los resultados puedan ser aplicados a la población de referencia. Una de las estrategias para lograr validez externa consiste en realizar muestreos probabilísticos o aleatorios, en los que todos los individuos que conforman la población de estudio tengan una probabilidad conocida y diferente de cero para ser incluidos en la investigación.

Otros tipos de error, que pueden ser frecuentes durante la realización de estudios transversales cuando no se aplican mecanismos de control adecuados, son los siguientes: a) el de cortesía, que se presenta cuando la persona investigada trata de complacer al entrevistador proporcionándole la respuesta que considera más adecuada o socialmente aceptable;[10] b) el de vigilancia, que aparece cuando existen condiciones que permiten una mejor vigilancia (clínica, de laboratorio o epidemiológica) de la enfermedad entre la población de estudio que en la población general, y c) el de información, que resulta de la obtención de datos poco verídicos o incompletos, o de la falta de respuesta de los individuos seleccionados para el estudio.

Con el fin de analizar y corregir el efecto de la falta de participación, es necesario conocer las razones de no respuesta y otras características de los sujetos no participantes, para conocer si se trata de valores distribuidos al azar o de manera sistemática y estimar de qué manera podrían afectar los resultados del estudio. La comparación entre los estimadores obtenidos con y sin datos de estos sujetos nos permitirá estimar la magnitud del error o del sesgo, según sea el caso. La inclusión de los sujetos no participantes permite simular diferentes escenarios, incluyendo el estimador más conservador, o sea, el del peor escenario, sesgado contra la prevalencia o hipótesis de interés. Durante la etapa de captura de información es importante también asegurarse de que no se pierdan datos; frecuentemente puede haber omisiones o datos ilegibles en el momento de la captura, lo que ocasionará la ausencia o pérdida de cierto número de datos. Será recomendable incluir funciones de revisión de dispersión de valores posibles durante la digitación de información y desarrollar esquemas de doble entrada de información para tener una vigilancia de la calidad de este proceso y minimizar los errores. La tabulación y el análisis de variables con va-

lores perdidos puede cambiar de manera importante el número de observaciones que se incluye en dichos análisis, lo que también puede afectar la validez del estudio.

Cuando se hacen preguntas sobre exposiciones o eventos pasados, las personas que han atravesado por experiencias que tienden a ser recordadas con mayor facilidad (por ejemplo, las experiencias traumáticas, como enfermedades, abortos o accidentes) pueden recordar las exposiciones con mayor precisión que los que no tuvieron experiencias similares; lo anterior da lugar a la aparición de un error distribuido sistemáticamente en los individuos de la muestra, que se conoce como sesgo de memoria.

En resumen, puede señalarse que la ausencia de sesgos en la selección de los sujetos de estudio y en la medición de variables en la población de estudio garantiza su validez interna; esto es, los resultados obtenidos son ciertos para la población o muestra estudiada. Si la muestra es representativa de la población base aumentará la validez externa del estudio; esto es, la posibilidad de extrapolar dichos resultados a la población blanco o población de la cual se obtuvo la muestra.

Por lo general, el tamaño de muestra en las encuestas transversales se calcula de tal forma que permita estimar, con un determinado poder y nivel de confianza, la prevalencia de alguna enfermedad o característica de la población, o bien diferencias en nuestra variable de resultado (daño, enfermedad o evento) de acuerdo con la variable de exposición. Un tamaño de muestra pequeño no permitirá que el estudio tenga el poder suficiente para encontrar asociaciones significativas entre las variables de exposición y el resultado; por el contrario, un tamaño excesivo ocasionará la aparición de significancia estadística de estimadores de poca magnitud, que indican diferencias de poca relevancia biológica y cuya búsqueda, por el contrario, implica dispendio de recursos y tiempo.[11,12]



## DEFINICIÓN DE VARIABLES EN ESTUDIOS TRANSVERSALES

Antes de iniciar el desarrollo de una encuesta transversal, es importante definir las variables de interés y, dentro de éstas, especificar cuáles serán las variables independientes, las de exposición y las que potencialmente pueden jugar un papel como variables confusoras o modificadores de efecto que desean estudiarse. Esta definición, como en todos los diseños epidemiológicos, debe ser teórica y operacional. La definición teórica indica el sitio en donde la variable se encuentra ubicada dentro del modelo que nos sirve de marco conceptual y en relación con el resto de variables a investigar; la operacional consiste en determinar la manera concreta en que se observará y medirá una variable. Ésta, junto con los indicadores e instrumentos que se utilizarán, definirá el tipo de análisis de dichas variables (fig. 1 y cuadro I).

## CONDUCCIÓN DE ENCUESTAS TRANSVERSALES

Las encuestas transversales, si bien son logísticamente más sencillas que los estudios que implican un seguimiento de individuos, tienen dificultades importantes por los tamaños de muestra que deben alcanzarse y porque se requiere de un marco muestral apropiado. Una vez diseñado el estudio, definidas la población y muestra, así como las variables que se investigarán, es necesario definir los instrumentos que serán empleados para recolectar la infor-

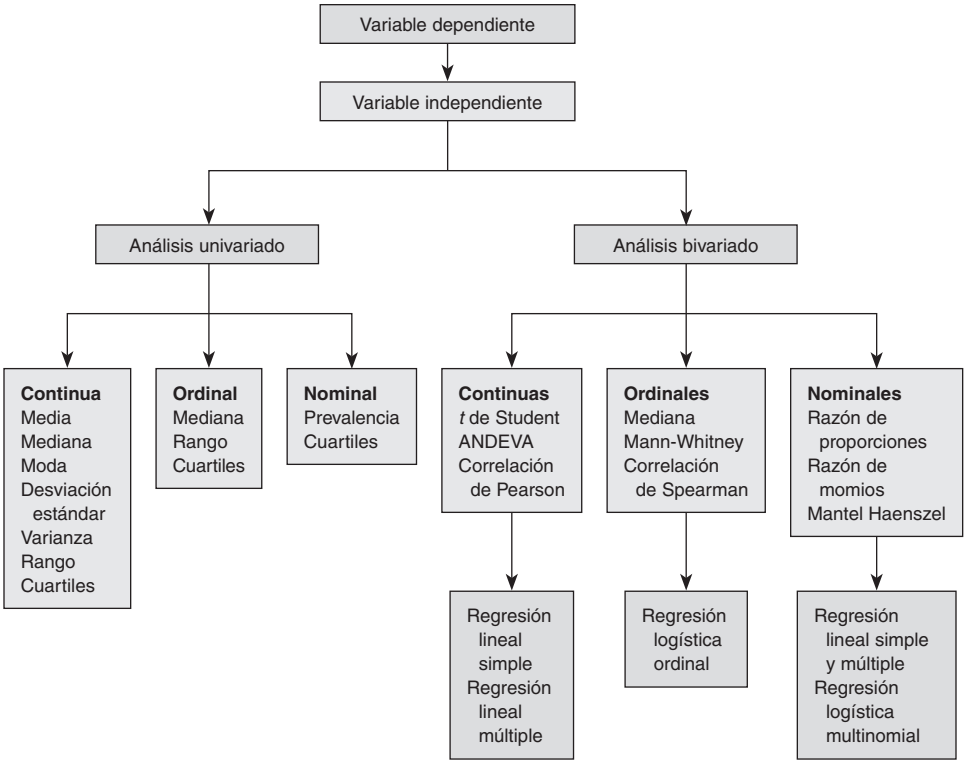



**Figura 1 Análisis estadístico por tipo de variable**

mación. Las encuestas transversales con frecuencia utilizan cuestionarios que se pueden aplicar a los informantes por medio de un entrevistador, o cuestionarios autoadministrados; en otras ocasiones se tomarán muestras biológicas (sangre, orina, saliva), se harán mediciones antropométricas (peso, talla, pliegues cutáneos) o se consignará información del hogar (disponibilidad de agua potable, manejo de excretas, etc.).

En el caso de la utilización de cuestionarios, es necesario definir al informante ideal que pueda proporcionar la información necesaria para el estudio. En general, la persona que responde debe ser capaz de entender el vocabulario utilizado en las preguntas. Los datos de ingreso económico familiar deberá proporcionarlos el jefe o jefa de familia. Asimismo, es necesario definir si los cuestionarios serán administrados por una entrevistadora o entrevistador (ya sea cara a cara o vía telefónica), o si los cuestionarios serán contestados por escrito por los informantes (ya sea en presencia del personal del proyecto o por otros medios como el correo). Ante el avance de los recursos computacionales, también es posible que la información se capture en medios electrónicos en el momento de la entre-

## Cuadro I Ejemplo de sesgo de selección

Un cardiólogo desea estudiar a pacientes en alto riesgo de enfermedad coronaria, por lo que solicita a los médicos de consulta externa que recluten pacientes con hipertensión arterial. Los médicos de consulta externa identifican a 3 000 pacientes con diagnóstico de hipertensión arterial y los invitan a participar en el estudio. Por diversos motivos, sólo 300 pacientes acceden a ser hospitalizados por un día.

Con el fin de estudiar la relación entre obesidad y cardiopatía coronaria (CC), el cardiólogo clasifica a los 300 pacientes de la manera siguiente:

| | Obesos | No obesos | Total |
|---|---|---|---|
| CC | 10 | 10 | 50 |
| No CC | 20 | 230 | 250 |
| Total | 30 | 270 | 300 |

RM = 2.87; IC95% 1.27-6.5; *x2*(1) = 6.67 *p* = <0.05

Como puede verse, la posibilidad (momios de prevalencia) de CC en los sujetos obesos es casi tres veces mayor (RM = 2.87) que en los no obesos. Esta relación es estadísticamente significativa a un alfa de 0.05 y el IC a 95% muestra valores de RM que excluyen el valor nulo. Posteriormente, el cardiólogo, de alguna manera, logra estudiar a la población de 3 000 pacientes hipertensos y encuentra lo siguiente:



| | Obesos | No obesos | Total |
|---|---|---|---|
| CC | 25 | 175 | 200 |
| No CC | 250 | 2550 | 2800 |
| Total | 275 | 2725 | 3000 |

RM = 1.45; IC95% 0.94-2.25; *x2*(1) = 2.86 *p* = >0.05

Para sorpresa del investigador, en el segundo estudio, la RM disminuye casi a la mitad, y desaparece la significancia estadística encontrada previamente. Además, el IC incluye el valor nulo de no asociación. Esta discordancia se debe a que en el estudio inicial las proporciones de participantes en los cuatro subgrupos difieren de las proporciones de los mismos subgrupos en la población total de 3 000 pacientes, lo que sugiere un sesgo de selección.

Comparando las tablas, nótese que:

- de los 25 pacientes obesos con CC, 10 participaron en el primer estudio (40%)
- de los 250 pacientes obesos sin CC, 20 participaron en el primer estudio (8%)
- de los 175 pacientes no obesos con CC, 40 participaron en el primer estudio (23%)
- de los 2 550 pacientes no obesidad sin CC, 230 participaron en el primer estudio (9%)

vista. Esta decisión dependerá de las características de los informantes y de la naturaleza de la información a recolectar. Las preguntas sobre conducta sexual, toxicomanías, y otras sobre la vida privada de las personas, suelen producir respuestas incompletas o evasivas por parte de los participantes, por lo que deberán probarse de antemano y hacerse con particular cuidado mediante entrevistadores debidamente calificados y entrenados, en condiciones que aseguren la confidencialidad y tranquilidad de los participantes.

Es necesario cuidar la integración de los instrumentos de recolección de información. Los cuestionarios deben adaptarse a la forma en que serán administrados y a la población bajo estudio. Los cuestionarios deben tener un formato que permita su aplicación y, posteriormente, su fácil codificación (transcripción de respuestas a códigos numéricos) y captura de información en medios electrónicos. Debe evaluarse la confiabilidad y validez de un cuestionario antes de utilizarlo en cualquier estudio. La confiabilidad se refiere a la capacidad de un instrumento para dar resultados similares en distintos momentos sobre características que se consideran invariantes en el periodo comprendido entre dichos momentos. La validez de un instrumento se refiere a su utilidad para medir realmente la característica o concepto que desea medir.[13] La confiabilidad y validez de las distintas secciones de un cuestionario pueden evaluarse mediante un estudio específico (denominado estudio piloto o de validación) realizado con una muestra menor de sujetos.

Al integrar un cuestionario no es necesario que todas las preguntas que contiene se hayan elaborado por primera vez para el estudio en el que se emplearán. Es posible integrar en un cuestionario preguntas o escalas que ya se han utilizado y validado con esa población o con poblaciones similares. El uso de instrumentos de medición similares y estandarizados es deseable ya que permitirá comparar los resultados con los de otros estudios. En todos los casos es necesario llevar a cabo una prueba piloto de los instrumentos de recolección de información. Dicha prueba, llevada a cabo con una submuestra de la población bajo estudio, permitirá corregir errores y problemas en el cuestionario y su procedimiento de aplicación. Asimismo, si en dicha submuestra es posible obtener y verificar las respuestas correctas al cuestionario o instrumento de medición, podrá estimarse el error existente entre las respuestas del cuestionario inicial (cuestionario a evaluar) y las del cuestionario validado (estándar de oro).[14] La estimación del error puede usarse a posteriori para corregir los resultados obtenidos con el instrumento de campo.

El trabajo de campo es una etapa crucial del estudio. Es en ese momento cuando se recopilará la información para los análisis, y los errores u omisiones que se produzcan en esta etapa serán difíciles de corregir. En consecuencia, es importante llevar a cabo un riguroso entrenamiento y supervisión del personal de campo. Habitualmente, el equipo de trabajo de campo está integrado por supervisores que revisan y coordinan la recolección de información, y entrevistadores y recolectores que propiamente recogen la información (ya sea administrando cuestionarios, tomando muestras biológicas o antropométricas, etc.). El personal de campo debe tener un conocimiento general del proyecto que le permita recolectar la información necesaria y minimizar la introducción de errores. El entrevistador o recolector de datos también puede ser fuente de error cuando un mismo entrevistador obtiene mediciones diferentes de la característica o atributo de interés (variabilidad intra observador), o cuando una misma medición se obtiene de manera diferente entre un observador y otro (variabilidad entre observadores). El entrevistador puede también puede ser fuente impor-

tante de sesgo, en especial cuando conoce la hipótesis de trabajo y esta información condiciona la aparición de diferencias en la manera de obtener la información entre los sujetos que tienen el evento en estudio y el resto de los participantes.

El personal de campo debe conocer a fondo los instrumentos de recolección de la información y el equipo a utilizar para la obtención de muestras; debe, asimismo, estar familiarizado con la toma de las muestras a recolectar en el estudio. Del mismo modo, debe conocer el esquema de muestreo a utilizar y los procedimientos a seguir cuando un informante no se encuentre, cuando una entrevista no pueda realizarse, cuando un participante potencial decline la participación, etcétera.

Una vez recolectada la información, los supervisores de campo y el personal de verificación deben revisarla cuidadosamente. En caso de detectar omisiones o anomalías, es recomendable regresar con la persona entrevistada para completar o corregir la información, la cual, una vez verificada, debe codificarse para que pueda ser capturada en medios electrónicos y, posteriormente, analizada. El sesgo del observador o el llenado fraudulento de datos pueden evaluarse mediante la identificación de repetición de ciertos dígitos en una variable registrada por un observador en comparación con los demás observadores. La distribución de respuestas a dicha variable deberá ser similar a la obtenida por los otros entrevistadores.

Para minimizar errores en este proceso es necesario emplear programas de captura validada, en los cuales la información es capturada dos veces para identificar discrepancias. Son comunes dos tipos de verificación: por rangos (no debe haber valores fuera del límite de los valores posibles de esa variable) y lógica (no debe haber valores incoherentes, como abortos en el sexo masculino, o profesionistas analfabetas). Actualmente es posible llevarlas a cabo en los programas computacionales de bases de datos relacionales (DBase, Paradox, FoxPro) y en algunos paquetes estadísticos que contienen rutinas para la elaboración de programas de captura validada de fácil utilización (Epi-Info, CDC, Atlanta). El uso de tecnologías, como lectores ópticos, puede agilizar el proceso de captura, aunque requiere de equipo y formatos de captura de información muy específicos.

El diseño, métodos y procedimientos del estudio deberán estar debidamente documentados, de tal manera que exista información disponible en caso de que sea necesario replicar y comparar el estudio.



## ANÁLISIS DE LAS ENCUESTAS TRANSVERSALES

El análisis de datos de una encuesta transversal depende de sus objetivos y del tipo de variables del estudio. Si el fin es caracterizar o describir la población, se miden las variables una vez y se presentan los valores de cada una de ellas o por grupos. El análisis de información de encuestas transversales suele iniciar con la obtención de estadísticas descriptivas de variables de interés (figura 1). Este análisis permitirá conocer las características generales de la población bajo estudio, y estimar prevalencias o promedios de las exposiciones y variables del resultado. Por ejemplo, una encuesta permitirá conocer la frecuencia y distribución de edades, escolaridad, ingreso económico, género, uso de servicios de salud, motivos de consulta médica, tabaquismo, cefalea, opinión sobre el estado de salud, etcétera. Para datos categóricos

(presencia o ausencia de enfermedad, número de hijos, nivel socioeconómico bajo, medio, alto, etc.) la descripción se hace por medio de la distribución de frecuencias (número de sujetos u observaciones dentro de cada categoría de la variable), frecuencias relativas (distribución porcentual de las observaciones dentro de las categorías de la variable) y proporciones. La prevalencia global de una enfermedad se obtiene dividiendo el número de casos encontrados al momento del estudio entre el total de la población de estudio.

Si quiere estimarse la prevalencia de una enfermedad, por ejemplo, de diabetes mellitus, los datos se presentan como proporción (número de diabéticos sobre el total de la población) por cien, mil, diez mil, etcétera, aunque es más frecuente presentar los datos en términos de porcentajes. Si la medición se realiza en un periodo muy corto se le llama prevalencia puntual; si se mide en un periodo mayor se le llama prevalencia lápsica. La prevalencia lápsica es una medida que combina incidencia y prevalencia, ya que para construirla se utilizan tanto los casos prevalentes como los casos nuevos (incidentes) presentados en el periodo de estudio.

En el caso de variables continuas, como el peso y la talla de niños menores de cinco años, se presentan medidas de tendencia central (media, mediana, moda) y de dispersión (desviación estándar, varianza, percentiles). Este tipo de variables puede categorizarse en escalas, de acuerdo con el interés de la investigación, es decir, convertirse en variables ordinales (por ejemplo, bajo peso, normal y sobrepeso, o talla baja, normal y alta).

En los estudios transversales analíticos se explora la relación o asociación entre variables de exposición y de resultado (o evento). En el caso de variables dicotómicas, la clasificación se hace en cuatro grupos: enfermos expuestos, enfermos no expuestos, sanos expuestos y sanos no expuestos (figura 2).

La asociación entre la ocurrencia de una enfermedad (variable de resultado) y algunos factores de riesgo (variables independientes o de exposición) se hace a través del cálculo de medidas de asociación. Como se señaló, la mejor medida del riesgo de pasar del estado sano al de enfermo en una población es la tasa de incidencia (cuyo numerador es la cantidad de casos nuevos de enfermedad y el denominador es el tiempo-persona en riesgo); la razón



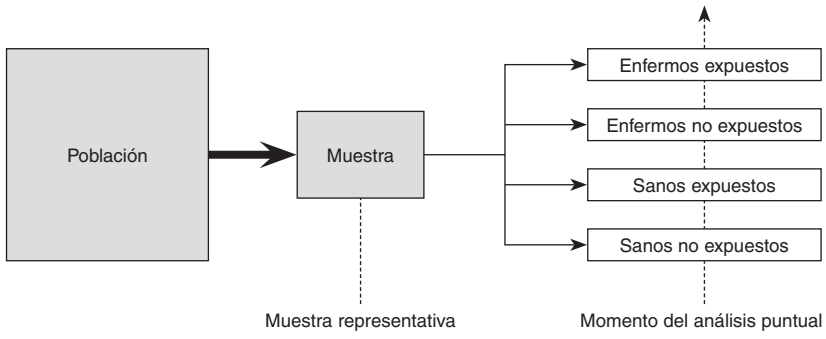



**Figura 2 Diseño de encuesta transversal**

de tasas (tasa de incidencia en expuestos entre la tasa de incidencia en no expuestos) cuantifica el exceso de riesgo de enfermedad entre expuestos y no expuestos. A falta de este estimador ideal, puede obtenerse la incidencia acumulada (número de casos nuevos de enfermedad entre la población de riesgo al inicio del estudio) y la razón de riesgos o riesgo relativo (incidencia acumulada en expuestos entre incidencia acumulada en no expuestos). En los estudios transversales únicamente es posible estimar la prevalencia [casos existentes al momento del estudio, entre grupos con ("expuestos") y sin ("no expuestos") cierta característica de la muestra poblacional], y por esta razón las únicas medidas de asociación que pueden obtenerse son la razón de prevalencias (RP) y la razón de momios de prevalencia (RMP).[15] Para la interpretación de estas medidas de asociación es necesario considerar la relación entre prevalencia, incidencia y duración de la enfermedad, la cual se expresa como:

$$P = I \cdot D \times (1 - P)$$

donde:
P = prevalencia
I = incidencia
D = duración promedio de la enfermedad

Para calcular estas medidas de asociación, se construye una tabla de cuatro celdas (cuando se trata de variables dicotómicas) en la cual las columnas registran el número de expuestos y no expuestos, y los renglones el número de enfermos y no enfermos:

| | Expuestos | No expuestos | Total |
|---|---|---|---|
| Enfermos | a | b | a+=b |
| No enfermos | c | d | c+d |
| Total | a+c | b+d | a+b+c+d |

donde:
$a + b$ = número de enfermos en la población
$a + b/a + b + c + d$ = prevalencia de enfermedad en la población
$a/a + c$ = prevalencia de enfermedad en los expuestos
$b/b + d$ = prevalencia de enfermedad en los no expuestos
$(a/a+c)/(b/b + d)$ = RP de enfermedad.
$(a * d)/(b * c)$ = RMP razón de momios de prevalencia

Un valor de 1 se interpreta como igual prevalencia de enfermedad entre expuestos y no expuestos. Un valor mayor de 1 significa que la prevalencia es mayor en los expuestos que en los no expuestos. Un valor menor a 1 significa que la prevalencia es menor en los expuestos que en los no expuestos.

La dependencia de la prevalencia respecto a la duración de la enfermedad, antes mencionada, hace que estas medidas se aproximen a la razón de riesgos sólo bajo ciertas condiciones. La RP se aproxima, o es buen estimador, de la razón de riesgos o riesgo relativo (RR) cuando la duración de la enfermedad es igual entre expuestos y no expuestos y cuando no hay migración hacia dentro o fuera de estos grupos.

$$RP = RR \times \frac{\text{D en expuestos}}{\text{D en no expuestos}} \times \frac{\left[\text{1.0 - prevalencia en expuestos}\right]}{\left[\text{1.0 - prevalencia en no expuestos}\right]}$$

donde:
D = duración de la enfermedad

Así, la RP puede ser un buen estimador del riesgo relativo en función de la razón de duración de la enfermedad en expuestos y no ex-

puestos, y de la razón de los complementos de la prevalencia en expuestos y no expuestos. Alternativamente, puede calcularse la razón de momios de prevalencia (RMP) de enfermedad, con la fórmula:

**$(a/c)/(b/d)$, o con la razón de productos cruzados $(a * d)/(b * c)$**

La RMP también se relaciona con la incidencia y duración de la enfermedad, ya que:

$$\text{Momios de prevalencia} = \frac{\text{prevalencia}}{\text{1-prevalencia}} =$$

$$= \text{incidencia} \times \text{duración}$$

Así, la RMP es igual a:

$$\frac{\text{Momios de prevalencia en expuestos}}{\text{Momios de prevalencia en no expuestos}}$$

El valor de la RMP es aproximado al de la RP y, en consecuencia, al valor del RR, cuando la prevalencia de la enfermedad es baja (menor que 10%)* y su interpretación es similar: un valor de 1 se interpreta como momios iguales o igual posibilidad de enfermar entre expuestos y no expuestos; un valor mayor de 1 significa que la posibilidad de enfermar es mayor en los expuestos que en los no expuestos, y un valor menor a 1 significa que la posibilidad de enfermar es menor en los no expuestos que en los expuestos. Sin embargo, la RMP es difícil de interpretar como una medida de efecto y es inconsistente con respecto a la medida deseable —que es la Razón de Tasas de Incidencia (RTI)— ya que la RMP puede sub o sobreestimar la RTI. Además, la RMP puede diferir de los resultados obtenidos mediante la RP en lo que respecta a la estimación de modificaciones de efecto y el control de variables confusoras.[16]

Por ejemplo, en un estudio transversal sobre prevalencia y factores de riesgo para infección por *Helicobacter pylori*,[17] los autores obtuvieron los siguientes estimadores:

**Razón de prevalencias y razón de momios para infección por H. pylori, ajustada por edad**

| Región | RP | RM |
|---|---|---|
| Rural | 1.0 | 1.0 |
| Semiurbana | 1.17 (1.09 − 1.26) | 1.32 (1.16 1.51) |
| Urbana | 1.24 (1.16 − 1.32) | 1.49 (1.32 1.68) |

Nótese cómo la RM siempre sobreestima la RP (y la razón de riesgos en los estudios de incidencia), lo que, además, aumenta la amplitud de los intervalos de confianza. Con estimadores cercanos al 1, la diferencia puede implicar la inclusión (si usa RP) o no (si usa RM) del valor nulo. La significancia estadística podría verse igualmente afectada. Estos datos indican la discordancia que puede ser observada entre ambas medidas. Como se mencionó anteriormente, lo recomendable es utilizar en la medida de lo posible las RP, ya que tienen una interpretación clara y directa y bajo ciertas suposiciones son un buen estimador de la RTI.

La significancia estadística de la asociación entre las variables se hace por medio del cálculo de intervalos de confianza y la prueba de hipótesis de no asociación, donde la hipótesis nula (o $H_0$) es que la RP o la RM son iguales a 1. Además de las medidas relativas de asociación (RP y RM), es de interés el cálculo de medidas absolutas tales como el riesgo atribuible, la fracción etiológica y la fracción prevenible. Para la interpretación de estas medidas,

* Los momios [a/c] y [b/d] son prácticamente iguales a las proporciones [a/a + c] y [b/b + d] cuando éstas son pequeñas, como en el caso de una prevalencia menor a 1 por ciento.

siempre deberá tenerse en mente la naturaleza transversal de los datos, en cuanto a que solamente aproximan el riesgo. Estas medidas tienen mayor aplicabilidad para la planeación de programas de prevención primaria y secundaria, ya que miden el impacto potencial de la eliminación de los factores de riesgo sobre la población:

- Riesgo atribuible: es la diferencia entre la prevalencia de enfermedad en el grupo expuesto y el no expuesto, o sea, la diferencia de prevalencias. Nótese que, por tratarse de estudios transversales, las prevalencias suelen ser mediciones aproximadas del riesgo. Se obtiene así:

  P expuestos − P no expuestos:
  a / (a + c) − b (b + d)

- Fracción etiológica: representa la proporción de casos de enfermedad entre los expuestos al factor de riesgo y que son atribuibles a la exposición a dicho factor:

  FE: RM + 1 / RM

- Fracción atribuible de la población: expresa la proporción de casos de enfermedad ocurridos en el total de la población (expuestos y no expuestos) que se atribuyen a la exposición al factor asociado (y por lo tanto la proporción o porcentaje prevenido si se elimina dicha exposición). Se obtiene con:

  FAP: RM − 1 / RM * a / a + b

- En el caso de factores que no son de riesgo, sino de protección, como en el caso de la inmunización mediante la aplicación de vacunas, la fracción atribuible se conoce como fracción evitada. Ésta expresa la proporción de casos que se evitarían si se implementara el factor protector y se obtiene con:

  1 − RM

Debido a que, frecuentemente, la selección de sujetos se obtiene de la población general, el control de variables confusoras suele hacerse durante el análisis de los datos (y no en el diseño). Para estimar el efecto de las variables confusoras (que, como se ha señalado antes, son aquéllas que se encuentran asociadas tanto con el efecto como con la exposición, pero sin que se encuentren en la trayectoria de la cadena causal que los liga), será siempre deseable, dado que estamos en el contexto de estudios observacionales, realizar ajustes directos o indirectos, análisis estratificado o técnicas de análisis multivariado —regresión lineal múltiple, regresión logística, etc. (figura 1)—. En el caso del análisis estratificado el ajuste por variables confusoras se hace con el método de Mantel-Haenszel. La RMP ajustada se interpreta de manera similar a la RMP cruda, pero se le refiere como ajustada por, o controlando el efecto de, una tercera o más variables.

En el caso de variables de resultado continuas y variables de exposición categóricas podrá hacerse una comparación de medias con pruebas como la de *t* de Student, análisis de varianza o modelos de regresión. También es posible utilizar métodos estadísticos no paramétricos, algunos de los cuales se incluyen en la figura 1 y en el cuadro II. En el apartado de análisis estadístico se describen con amplitud sus aplicaciones y su interpretación.

En el caso de variables continuas de resultado y de exposición, la relación lineal entre dos variables continuas puede llevarse a cabo con la estimación de la correlación de Pearson, o bien mediante el análisis de regresión lineal

**Cuadro II Método de análisis e interpretación por tipo de variables**

| *Variable dependiente* | *Variable independiente* | *Método* | *Interpretación* |
|---|---|---|---|
| Categórica | Categórica | *Chi* cuadrada | Diferencia estadísticamente significativa entre los grupos tabulados si $p < 0.05$ |
| Continua | Continua(s) | Correlación | Asociación lineal; 0 no asociación, −1 asociación negativa, +1 asociación positiva. |
| Continua | Categórica | Diferencia de medias, análisis de varianza | La media de la característica y de uno o más grupos es diferente o no a la media en el(los) otro(s) grupo(s) |
| Continua | Continua, ordinal [puede ser nominal (indicadora o *dummy*) o dicotómica] | Regresión lineal | Incremento* en el valor promedio de resultado y por cada unidad de aumento en x1, ajustando por todas las demás variables del modelo |
| Dicotómica | Continua, ordinal | Regresión logística | Incremento* promedio en momios de resultado y por cada unidad de aumento en x1, ajustando por todas las demás variables de modelo |
| Dicotómica | nominal (indicadora o dummy) | | Incremento* promedio en momios de resultado y para categoría x1, comparada con categoría de referencia, ajustando por todas las demás variables de modelo |
| Dicotómica | dicotómica | | Incremento promedio en momios de resultado y para categoría x1, comparado con categoría de referencia, ajustando por todas las demás variables de modelo |
| Ordinal | Continua, ordinal, nominal (indicadora o dummy), dicotómica | Regresión logística ordinal | Incremento promedio en los momios de la variable y por cada unidad de aumento en x1, ordenada ordinalmente, ajustando por todas las demás variables de modelo |
| Multinomial | Continua, ordinal [puede ser nominal (indicadora o *dummy*) o dicotómica] | Regresión logística multinomial | Razón entre el riesgo relativo de categoría y, (comparado con categoría de referencia y), por unidad de incremento en x1 y el riesgo relativo cuando x no cambia, ajustando por todas las demás variables de modelo. |

* Coeficientes positivos

simple. Este último método permite establecer una relación lineal entre las variables de exposición (variable predictora o independiente $x_i$) y de resultado (o dependiente $y$), utilizando la fórmula general $y = a + b_i x_i ...$, donde $a$ y $b$ son las constantes estimadas a partir de los datos.[18] Para el ajuste de variables confusoras o para evaluar modificación del efecto o interacción, es posible también emplear técnicas de análisis multivariado, ya sea usando regresión lineal múltiple (para variables de exposición categóricas y continuas y variables de resultado continuas) o regresión logística o binomial (para variables de resultado dicotómicas) y ordinal o multinomial (para variables ordinales y nominales con más de dos categorías) (figura 1 y cuadro II). Es importante mencionar que la construcción de modelos multivariados —la selección de variables a incluir en el modelo— para el análisis de encuestas transversales, debe guiarse por la teoría, por el tipo de muestreo y por análisis cuidadosos de bondad de ajuste y del efecto de cada variable en el modelo. Por esta razón, los métodos de regresión gradual (*stepwise)* anterógrada o retrógrada, los cuales se basan únicamente en el incremento del valor explicativo del modelo, deben utilizarse con sumo cuidado.

La regresión logística permite obtener RMP crudas y ajustadas por potenciales confusores, mientras que la regresión binomial o de riesgos proporcionales permite estimar RP crudas y ajustadas. Para estos estimadores crudos y ajustados es posible realizar pruebas de hipótesis de igualdad del valor nulo, calcular intervalos de confianza, así como realizar pruebas de bondad de ajuste. Uno de los métodos para el cálculo del IC de la RMP más utilizado es el de Woolf:

$$\text{IC 95\% RM} = \\ = (\text{Ln RM}) \pm Z\sqrt{1/a + 1/b + 1/c + 1/d}$$

Ejemplo de análisis estratificado:
El siguiente ejemplo se basa en datos de Camargo[19] y Lazcano:[20]

**Asociación entre ingestión de bebidas alcohólicas y seropositividad a *H. pylori***

| | Expuestos | No expuestos | Total |
|---|---|---|---|
| Casos | 1171 | 1616 | 2787 |
| No casos | 1043 | 2031 | 3074 |
| Total | 2214 | 3647 | 5861 |

RMP = 1.41; IC95% 1.26 − 1.57
RP = 1.19; IC95% 1.13 − 1.26
Valor $p < 0.01$

Suponga que desea saber si el ser mujer u hombre es una variable que influye sobre la asociación entre la ingestión de alcohol y la positividad a *H. pylori*:

Para las mujeres:

**Asociación entre ingestión de bebidas alcohólicas y seropositividad a *H. pylori***

| | Expuestos | No expuestos | Total |
|---|---|---|---|
| Casos | 626 | 1050 | 1676 |
| No casos | 595 | 1294 | 1889 |
| Total | 1221 | 2344 | 3565 |

RMP = 1.29; IC95% 1.12-1.49
RP = 1.14; IC95% 1.06-1.22
Valor $p < 0.01$

La RMP y la RP específicas para el estrato muestran una asociación positiva entre ingestión de alcohol e infección por *H. pylori*; nótese que los valores son similares al valor crudo.

Para los hombres:

Asociación entre ingestión de bebidas alcohólicas y seropositividad a *H. pylori*

| | Expuestos | No expuestos | Total |
|---|---|---|---|
| Casos | 545 | 566 | 1111 |
| No casos | 448 | 737 | 1185 |
| Total | 993 | 1303 | 2296 |

RMP = 1.58; IC95% 1.33-1.87
RP = 1.26; IC95% 1.16-1.37
Valor p $< 0.01$

La RMP y la RP específicas para el estrato muestran una asociación positiva entre ingestión de alcohol e infección por *H. pylori*; nótese que los valores son similares al valor crudo.

Si los estimadores específicos por estrato fueran diferentes en magnitud (con o sin significancia estadística) sospecharíamos la presencia de una interacción o modificación de efecto, fenómeno importante que debe notificarse cuando sea biológicamente plausible.[19] En este caso, la similitud entre los valores de las RM y RP específicos por estrato sugieren que no existe un efecto modificador (o interacción) del sexo sobre la asociación entre la ingestión de alcohol y la positividad a *H. pylori*. Las diferencias entre los estimadores específicos de los estratos pueden someterse a prueba de significancia estadística de homogeneidad de los estratos, usando el valor *p* convencional $\leq 0.05$.

Después de comprobar que la variable sexo no tiene un efecto modificador o de interacción, se procedería a evaluar el potencial efecto confusor del sexo sobre la asociación entre la ingestión de alcohol y la positividad a *H. pylori*. Una diferencia entre los valores de las RM y RP específicos por estrato y los valores crudos sugeriría que el sexo es un confusor potencial; sin embargo, observamos que dichos valores son muy similares, por lo que el sexo no actúa en este caso como confusor de la asociación entre la ingestión de alcohol y la positividad a *H. pylori*. Si se sospechara confusión, el estimador ajustado para la RMP se obtiene con la fórmula de Mantel-Haenszel (que comúnmente se aplica para incidencia acumulada):

$$OR \text{ de } M-H = \frac{\sum \frac{aixdi}{Ni}}{\sum \frac{bixci}{Ni}}$$

RM M-H

donde la "i" representa cada uno de los estratos y N el tamaño total de cada estrato, con lo que se obtiene:

RMP (M − H) = 1.40; IC95% 1.26-1.56

Esta RMP es muy similar a la RMP cruda (RM = 1.26; IC95% 1.26-1.57), lo que evidencia nuevamente que el sexo no es un confusor de la asociación entre ingestión de alcohol y positividad a *H. pylori*.

La sobrestimación de las razones de momios en comparación con las razones de prevalencias puede verse también por medio de los estratos. En este caso nos conduce a las mismas conclusiones, pero puede darse una situación donde el IC incluya o no el valor nulo (y el valor de *p* sea significativo o no), dependiendo del estimador obtenido (RMP o RP). La *chi* cuadrada indica diferencia estadísticamente significativa.

El análisis estratificado puede reproducirse con modelos de regresión logística, tanto para evaluar la interacción como la confusión. Esta última se ilustra a continuación:

Modelo de regresión logística para asociación entre positividad a *H. pylori* e ingestión de alcohol, ajustado por sexo:

RMP (alcohol) = 1.40; IC95% 1.26-1.56,
$p = 0.000$

RMP (sexo) = 1.02; IC95% .92-1.13,
$p = 0.46$

Modelo de regresión binomial para asociación entre positividad a *H. pylori* e ingestión de alcohol, ajustado por sexo:

RP (alcohol) = 1.19; IC95% 1.12-1.25,
$p = 0.000$

RP (sexo) = 1.01; IC95% 0.96-1.07,
$p = 0.54$

El análisis de regresión nos aporta además la RM o RP para la variable sexo, ajustada por ingesta de alcohol. El valor *p* convencionalmente aceptado para rechazar una hipótesis nula es de 0.05 o menos, y se define como la probabilidad de observar un valor tan extremo o más como el observado, bajo la suposición de que la hipótesis nula es cierta. Debe considerarse que el valor de *p* es una función del estimador a obtener y del tamaño de la muestra. Con un valor de confianza generalmente de 95%, los intervalos de confianza nos dan valores mínimos y máximos del estimador obtenido (prevalencia, RP, RM, diferencia de medias, correlación de Pearson, coeficiente beta de la regresión lineal); y se interpretan como 95% de confianza de que el parámetro verdadero del estimador se encuentre entre el límite inferior y el límite máximo. Si el intervalo de confianza no incluye el valor nulo del estimador dado, existe asociación estadística y el valor *p* será significativo a un alfa menor de 0.05. La amplitud del intervalo es función además del tamaño muestral. La información que proporcionan el valor *p* y los intervalos de confianza es complementaria, aunque el valor límite de 0.05 para *p* es arbitrario y el IC más informativo; una *p* mayor a 0.05 en presencia de un IC para una RMP de 0.7 a 20 significa que, aunque no tenemos una *p* estadísticamente significativa, la RMP podría ser hasta de 20 si se tuviera un tamaño de muestra mayor (asumiendo el mismo error de medición y ausencia de sesgos).

Las pruebas de bondad de ajuste evalúan la capacidad probabilística del método estadístico para ajustar o explicar la relación y variabilidad entre las variables incluidas en el modelo explicativo de exposición y efectos. Su significancia estadística se evalúa también con valores de *p*.

Como se mencionó, en ocasiones una muestra no refleja la composición de la población de la cual fue extraída directamente. Algunos grupos pueden estar sub o sobrerrepresentados; esto tiene efectos negativos sobre la validez interna y externa de los estimadores. Un ejemplo de este problema es el sesgo de selección, que se ilustra en el cuadro I.[11]

En el ejemplo del cuadro I es notorio que en el estudio participó una mayor proporción de pacientes obesos con cardiopatía coronaria (CC). Esta falta de representatividad de la muestra produjo una asociación espuria (falsa) entre la obesidad y la CC, derivada de un sesgo de selección. La selección aleatoria, característica de una encuesta bien realizada, habría tendido a obtener proporciones de sujetos participantes en cada celda similares a las de la población base, lo cual habría evitado este problema. Además nótese que, a pesar del aumento sustancial del tamaño muestral, se perdió la significancia estadística, lo que contradice la noción común de que, al aumentar el tamaño muestral, "seguramente" se logrará significancia estadística.

Las encuestas transversales frecuentemente exploran múltiples exposiciones y varia-

bles. Esto implica que, durante el análisis, deben construirse múltiples subgrupos. Con cada nueva hipótesis que se analiza aumenta la probabilidad de encontrar al azar asociaciones estadísticamente significativas. Existen métodos (por ejemplo el de Bonferroni) para tratar de corregir dicho problema; sin embargo, la naturaleza de los estudios transversales de por sí se limita a la exploración y generación de hipótesis.

El análisis de subgrupos dentro de un estudio transversal puede reducir mucho el número de individuos analizados, con la consiguiente pérdida de poder, sobre todo si se recuerda que la muestra se obtiene con base en una hipótesis en el grupo total. Su correcta interpretación deberá ser en el sentido de falta de significancia estadística al nivel convencional por falta de poder, y no de ausencia de efecto o de asociación, además de la consideración de sesgos.

En encuestas con grandes muestras (como las encuestas nacionales donde se obtienen datos sobre miles de individuos) es fácil encontrar asociaciones estadísticamente significativas; sin embargo, deberá considerarse la significancia conceptual (etiológica o biológica) de efectos pequeños, aunque estadísticamente significativos, y de efectos grandes, estadísticamente no significativos.

## TIPOS DE ENCUESTAS TRANSVERSALES

Naturalmente, existen muchos tipos de encuestas transversales. Las más comunes son las utilizadas para investigar la existencia de posibles relaciones entre determinadas condiciones presentes en una población y otra condición —generalmente considerada un daño— que se sabe o se sospecha presente en ella. No obstante, las encuestas transversales se usan frecuentemente para identificar y medir la frecuencia y distribución de fenómenos poblacionales que no son daños a la salud, pero que resultan importantes para conocer las necesidades sanitarias de la población, el uso de los recursos para la salud y la forma en que ambos aspectos se combinan. Asimismo, es común el uso de encuestas transversales para conocer la prevalencia de ciertas características en la población que resultan importantes para el diseño de los programas y servicios de salud, pero sobre las cuales existe información incompleta o limitada, que además no puede recogerse por medio de los sistemas rutinarios de información. Estas características pueden referirse a condiciones clasificadas como daños, a condiciones consideradas como riesgos o a aquéllas que indican la presencia o que modifican el curso de cualquiera de los dos elementos anteriores. Los organismos internacionales —la Organización Mundial de la Salud y el Fondo de las Naciones Unidas para la Infancia, por ejemplo— utilizan cada vez con mayor frecuencia este tipo de diseño, para identificar las tendencias mundiales de aquellas condiciones relacionadas con la salud que no recogen los sistemas tradicionales de información.

Por su relativa sencillez, las encuestas transversales se usan muy frecuentemente, y desde hace algunas décadas existe una tendencia mundial, cada vez más importante, orientada a la realización de estudios nacionales, diseñados en forma de encuestas transversales. En México se han llevado a cabo varias encuestas de salud, con representatividad a escala estatal y nacional. Éste es el caso de las encuestas nacionales de salud, la de enfermedades crónicas, las de seroprevalencia, las de adicciones y las de nutrición.[21-23]

## CONSIDERACIONES ÉTICAS Y DE BIOSEGURIDAD

El participante en una encuesta deberá estar enterado y de acuerdo con el uso que se le dará a la información que proporcione. Deberá garantizarse la seguridad, confidencialidad y, de ser posible, el anonimato de la persona que proporciona los datos. Debe evitarse el uso de datos para fines diferentes a los que autorizó el sujeto del estudio. Las muestras biológicas de sangre, material genético, tejidos y otras, deberán utilizarse exclusivamente para los fines autorizados por el participante y los residuos biológicos deben manejarse de manera adecuada con apego a las normas establecidas para ello. El uso de este material con objetivos de investigación distintos a los autorizados, aun años después del almacenamiento, requiere del consentimiento del donador.

Por último, es responsabilidad del investigador asegurarse de la calidad de los datos, tanto de los que obtuvo por medio de entrevistas o cuestionarios como de los correspondientes a mediciones de laboratorio, mediante sistemas de control de calidad. Una vez recolectados los datos, su manejo, análisis e interpretación deben realizarse de acuerdo con el protocolo de estudio y deberá evitarse la manipulación de los mismos hasta obtener resultados "interesantes" o convenientes.

## CONCLUSIONES

Las encuestas transversales son un diseño de investigación ampliamente utilizado. Entre sus ventajas podemos mencionar su bajo costo y rapidez, ya que no requieren del seguimiento de los sujetos de estudio. Este diseño permite explorar múltiples exposiciones y efectos, generar hipótesis y datos útiles para la planeación y gerencia de los servicios de salud, así como realizar mediciones de carga de la enfermedad. No obstante, este diseño también tiene importantes limitaciones como son: la imposibilidad de establecer causalidad; la falta de temporalidad de la asociación exposición-efecto (salvo en algunos casos en los que, por razones teóricas, es obvio que la exposición antecede a la variable de resultado); la dificultad para establecer valores basales para su comparación entre poblaciones y periodos; y su limitada utilidad para estudiar enfermedades de corta duración o poco frecuentes.



## REFERENCIAS


1. Kleinbaum DG, Sulliva KM, Barker ND. ActivEpi. Companion Book. A supplement for use with the ActivEpi CD-ROM. Nueva York: Springer-Verlag, 2003.
2. Gordis L. Epidemiology. Philadelphia: WB Saunders Co., 1996.
3. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Co., 1987.
4. Rothman KJ, Greenland S. Modern epidemiology. 2a. ed. Nueva York: Williams & Wilkins, 1998.
5. Dos Santos SI. Estudios transversales. En: Dos Santos I. Epidemiología del cáncer: principios y métodos. Lyon: Agencia Internacional de Investigación sobre el Cáncer/Organización Mundial de la Salud, 1999:225-244.
6. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. Monographs in epidemiology and biostatistics. Nueva York: Oxford University Press, 1996.

7. Walker AM. Observation and inference: An introduction to the methods of epidemiology. Boston: Epidemiology Resources, Inc., 1991.
8. Fowler FJ. Survey research methods. Applied social research methods series. 2a. ed. Vol. 1. Londres: Sage Publications, 1993.
9. Olaiz G, Rojas R, Barquera S, Shamah T, Aguilar C *et al*. Encuesta Nacional de Salud 2000. Tomo 2. La salud de los adultos. Cuernavaca: Instituto Nacional de Salud Pública, 2003.
10. Sackett DL. Bias in analytic research. J Chron Dis 1979;32:51.
11. Campbell D, Stanley J. Diseños experimentales y cuasi-experimentales en la investigación social. Buenos Aires: Amorrortu, 1982.
12. Altman D. Practical statistics for medical research. Londres: Chapman & Hall Ed., 1997.
13. Pagano M, Gauvreau K. Principles of biostatistics. Wadsworth, Belmont: Duxbury Press, 1993.
14. Babbie E. The practice of social research. 3ra. ed. Belmont: Wadsworth, Inc., 1983.
15. Aday L. Designing and conducting health surveys: A comprehensive guide. San Francisco: Jossey Bass, 1989.
16. Zhang JMB, Yu KF. What is a relative risk?: a method of correcting the odds ration in cohort studies of common outcomes. JAMA 1998;280(19):1690-1691.
17. Lilienfeld DE, Stolley PD. Foundations of epidemiology. 3ra. ed. Nueva York: Oxford University Press, 1994.
18. Selvin S. Statistical analysis of epidemiologic data. 2a. ed. Nueva York: Oxford University Press, 1996.
19. Camargo BC, Lazcano-Ponce E, Torres J, Velasco-Mondragón HE, Quiterio M, Correa P. Determinants of Helicobacter pylori seroprevalence in Mexican adolescents. Helicobacter 2004;9(2):106-114.
20. Lazcano-Ponce EC, Hernández-Prado B, Cruz VA, Allen B, Díaz R, Hernández GC *et al*. Chronic disease risk factors among healthy adolescents attending public schools in the state of Morelos, Mexico. Arch Med Res 2003;34 (3):222-236.
21. Instituto Nacional de Salud Pública. Encuesta Nacional de Salud II-1994. Cuernavaca, México: INSP, 1995.
22. Instituto Nacional de Salud Pública. ENSA 2000. Cuernavaca, México: INSP, 2001.
23. Dirección General de Epidemiología. Encuesta Nacional de Enfermedades Crónicas. México: Secretaría de Salud, 1994.

# Encuestas transversales — Ejercicios

Las preguntas de este ejercicio se basan en el siguiente artículo:

## RESUMEN

Parra-Cabrera S, Hernández-Ávila M, Tamayo-y-Orozco J, López-Carrillo L, Meneses-González F. Exercise and reproductive factors as predictors of bone density among osteoporotic women in Mexico City. Calcif Tissue Int 1996;59:89-94.

## INTRODUCCIÓN

Varios factores se han asociado a un incremento en la incidencia de fracturas por osteoporosis. Entre ellos se encuentran: falta de ejercicio, menopausia temprana, edad, uso de estrógenos, paridad, baja ingesta de calcio en la dieta, consumo alto en proteínas, consumo de cafeína y abuso de alcohol.

En México, la prevalencia de osteoporosis y sus complicaciones son inciertas. Un estudio reportó una tasa de mortalidad relacionada con osteoporosis de 0.06/100 000 para 1990. En un estudio más reciente, que incluyó una revisión de 5 000 certificados de defunción de mujeres de 50 años o más, se estimó una tasa de mortalidad de 1.8/1000.



### Material y métodos

*Población de estudio*

La población de estudio elegible correspondió a 378 mujeres mexicanas de 26 a 83 años de edad que acudieron, por primera vez, a una clínica de osteoporosis ubicada en la Ciudad de México, voluntariamente o referidas por su médico tratante, desde marzo de 1991 hasta febrero de 1992. Las mujeres correspondían al medio socioeconómico medio y alto. Firmaron una carta de consentimiento y la tasa de respuesta fue mayor a 90%. Se excluyeron a 65 mujeres por las siguientes razones: 36 no desearon participar en el estudio; 14 no reportaron la fecha de menopausia, y 15 reportaron una fractura secundaria a un traumatismo leve en los últimos 5 años.

*Recolección de información*

Las participantes contestaron un cuestionario autoaplicado en la clínica y siempre hubo un miembro del grupo de investigación para contestar cualquier pregunta. El cuestionario contenía preguntas sobre estilos de vida (p. ej. horas de ejercicio y de sueño, tabaquismo) y vida reproductiva. La información solicitada correspondía a las actividades realizadas en la semana que mejor representara las actividades del último año. Se consideró como menopausia natural a la suspensión de la menstruación, por lo menos durante el último año por razones no médicas, y como menopausia quirúrgica a la ocurrida mediante cirugía.

*Análisis estadístico*

Se utilizaron modelos de regresión lineal múltiple y coeficientes de correlación para medir la asociación entre las diferentes variables y la DMO tanto de columna lumbar como de la región fe-

moral. Se incluyeron las siguientes variables: edad, índice de masa corporal, índices de ejercicio y sedentarismo, paridad, menopausia, uso de estrógenos, y tabaquismo. El modelo final se construyó para cada variable resultado, mediante el método de regresión gradual retrógrada.

## RESULTADOS

Asumamos que los autores realizaron un análisis de regresión lineal simple inicial para predecir la DMO en la región femoral:

$$DMO = 0.934 + 0.0026\ X_1$$

donde:
$X_1$ = Indice de masa corporal (IMC)

A continuación se muestra un cuadro que contiene los resultados del análisis multivariado, obtenidos en mujeres con menopausia natural.

**Cuadro I Modelo multivariado[a] que indica la relación entre características específicas y densidad mineral ósea[b] de la región femoral**

| | Menopausia natural (n = 128) | | |
|---|---|---|---|
| *Variables* | β | *Error estándar* | *p* |
| Edad (años) | 0.001 | 0.0010 | <0.001 |
| IMC (kg/m$^2$)[c] | 0.009 | 0.0026 | <0.001 |
| Índice de actividad[d] | 0.008 | 0.0040 | 0.029 |
| Paridad (núm. de embarazos) | –0.002 | 0.0031 | 0.540 |
| Exfumador[e] | –0.015 | 0.0247 | 0.450 |
| Fumador actual[f] | –0.024 | 0.0239 | 0.290 |
| α | 0.934 | | |
| $R^{2g}$ | 0.310 | | |

[a] Regresión lineal múltiple
[b] Densidad mineral ósea en g/cm$^2$
[c] IMC = Índice de masa corporal
[d] Score de ejercicio. Transformación Log-e para mejorar normalidad.
[e] Exfumador: nunca fumador=0, exfumador =1, fumador actual = 0.
[f] Fumador actual: Nunca fumador = 0, ex-fumador = 0, fumador actual = 1.
[g] Proporción de la variabilidad total de la densidad mineral ósea explicada por las variables independientes.

## PREGUNTAS:

1. Con la información que proporcionan el título y la introducción del artículo, y considerando que la osteoporosis tiene una relación inversamente proporcional con la densidad mineral ósea (DMO), definida como el contenido mineral óseo medido en gramos por centímetro cuadrado de hueso ($g/cm^2$), plantee uno o más objetivos de investigación.
2. ¿Cuál es la exposición y cuál el evento de interés?
3. ¿Qué tipo de estudio epidemiológico propondría usted para lograr el objetivo del presente trabajo? Comente.
4. ¿Cuál considera usted que fue el tipo de muestreo utilizado para la selección de las mujeres?
5. ¿Cuál fue el motivo por el que los investigadores excluyeron a las 15 mujeres con antecedente de fractura? ¿De qué forma habrían afectado sus resultados si las hubieran incluido?
6. De acuerdo con la información presentada en este estudio, ¿Cuál considera usted que podría ser la población base? Explique qué tan bien representa la muestra seleccionada a la población base.
7. De acuerdo con la selección de la muestra ¿en qué tipo de sesgo pensaría usted? ¿Cómo podría haberlo evitado?
8. Suponiendo que el sesgo estuviese presente, ¿qué tipo de validez estaría comprometida?
9. En el presente trabajo el cuestionario fue autoaplicado. Ante esta situación, indique qué tipo de variabilidad (mediciones diferentes de una misma característica) sería importante considerar y evaluar. Explique de qué manera lo haría.
10. ¿Cuál es la unidad de observación? Explique por qué.
11. En el contexto del presente trabajo, explique qué entiende por regresión gradual retrógrada y qué aspectos recomendaría cuidar al utilizarla.
12. De acuerdo con los resultados del modelo de regresión lineal simple, ¿cuál sería la DMO estimada de una mujer con un IMC de 19 y cuál sería el de una con un IMC de 25?
13. Haciendo uso del cuadro I, interprete el valor $\beta$ correspondiente al IMC.
14. Los autores refieren en el artículo que la actividad física, la paridad y el IMC son determinantes importantes de la DMO en esta población. ¿Qué opina de esta aseveración?
15. ¿Cómo valora usted la validez externa del estudio y a quiénes se podrían extrapolar los resultados?
16. ¿Qué estudio propone usted para evaluar si el IMC y la actividad física preceden a la DMO?



## RESPUESTAS

1. El objetivo que los autores se plantearon en el estudio fue evaluar la relación entre ciertos factores reproductivos, obesidad y estilos de vida (i.e. tabaquismo y ejercicio) y la DMO en una muestra de mujeres posmenopáusicas de la Ciudad de México.
2. Los factores reproductivos, la obesidad y el estilo de vida corresponden a las variables de exposición. La DMO corresponde al evento de interés.
3. Los investigadores decidieron realizar un estudio transversal. Este tipo de estudio no es el ideal para evaluar causalidad; pero tiene la ventaja de ser más económico y más fácil de implementar. Un estudio de cohorte sería un mejor estudio, ya que los sujetos de investigación podrían clasificarse en diferentes categorías de exposición y medir la DMO en di-

ferentes puntos en el tiempo. Este tipo de estudio permitiría corregir el problema de temporalidad inherente a los estudios transversales.

4. Muestreo por conveniencia.
5. Las mujeres, después de una fractura, pudieron cambiar sus estilos de vida, particularmente en lo referente al ejercicio. Lo anterior puede conllevar a lo que se conoce como causalidad reversa, es decir, que la fractura pudo condicionar un cambio en los patrones de ejercicio. Incluirlas podría provocar una subestimación de la asociación real entre, por ejemplo, la falta de ejercicio y la DMO.
6. La población base podría corresponder a mujeres hispánicas de 26 a 83 años de edad de nivel socioeconómico medio y alto. La muestra seleccionada corresponde a mujeres que acudieron voluntariamente al estudio o que fueron referidas por su médico. Al tratarse de una muestra de mujeres autoseleccionadas, estas mujeres pueden tener mayor riesgo de osteoporosis o estar más expuestas a los factores de riesgo que la población base.
7. En un sesgo de selección que podría haberse evitado, por medio de un muestreo probabilístico o aleatorio.
8. La validez externa, ya que la muestra fue por conveniencia. La validez interna estaría comprometida muy probablemente. Sin embargo, en los estudios que incluyen mujeres que acuden a realizarse una prueba de tamizaje (p. ej. DMO), puede resultar que tanto las mujeres afectadas (p. ej. baja DMO) como las no afectadas (p. ej. DMO normal) tiendan a tener una mayor probabilidad de estar expuestas a factores de riesgo conocidos, pues es el conocimiento de estar probablemente en riesgo lo que las anima a participar. Lo anterior podría dar como resultado una medida de asociación relativa no sesgada (i.e., sesgo de selección compensado).
9. Es importante considerar la variabilidad intraobservador, debido a que una misma mujer pudo contestar de forma diferente la misma pregunta. Para evaluar la variabilidad intraobservador se aplica el cuestionario en más de una ocasión a la misma mujer para verificar la concordancia de las respuestas. El segundo cuestionario debe autoaplicarse después de un tiempo razonable, para que la señora no recuerde lo que contestó en el primero.
10. La unidad de estudio en este trabajo es la mujer, debido a que la información se recolectó de forma individual.
11. Los autores iniciaron con un modelo saturado, es decir, un modelo multivariado que incluía todas las variables investigadas. En este caso se encuentran el evento de interés (o variable dependiente), los factores de exposición y el resto de covariables (consideradas variables independientes). Enseguida fueron gradualmente eliminando una a una cada variable independiente, hasta que no pudiese excluirse ninguna sin disminuir en más de 10% la variabilidad explicada por el modelo. Este tipo de modelos —que se basan únicamente en el incremento de su valor explicativo— deben utilizarse con sumo cuidado, pues se corre el riesgo de eliminar variables teóricamente importantes. Una forma de evitarlo es no excluir aquellas variables que tengan una potencial relación biológica con la variable dependiente.
12. Las mujeres con mayor IMC tienen mayor DMO, como se observa a continuación.

$$\text{DMO} = 0.934 + 0.0026\,(19) = 0.983\ \text{g/cm}^2$$
$$\text{DMO} = 0.934 + 0.0026\,(25) = 0.999\ \text{g/cm}^2$$

13. Por cada incremento en una unidad de IMC, se observó un incremento promedio de 0.009 g/cm$^2$ de DMO, después de hacer los ajustes por todas las variables contenidas en el cuadro I.
14. En la discusión, los autores indican que sus resultados deben interpretarse con cautela. En los estudios transversales se realiza una sola medición de las exposiciones y del evento en un momento dado. Por lo anterior, en este estudio no es posible determinar en forma confiable si la actividad física y el IMC precedieron a la DMO.
15. Bajo el supuesto de que no hubiese ocurrido un sesgo de selección, los resultados de este trabajo podrían extrapolarse únicamente a las mujeres que acuden a la clínica de osteoporosis de donde fue seleccionada la muestra (mujeres de nivel socioeconómico medio y alto). Sin embargo, no existe razón biológica alguna por la cual estos resultados no puedan extrapolarse a mujeres con un nivel socioeconómico bajo.
16. Un estudio de cohorte sería un buen diseño, ya que, después de realizar una medición basal tanto de la variable dependiente (DMO) como de las independientes (IMC y actividad física), se haría un seguimiento de las mujeres para evaluar si efectivamente tanto la actividad física como el IMC determinan la DMO.

# IX Investigación de brotes

*Pablo Kuri Morales*
*Fernando Meneses González*
*Esteban Rodríguez Solís*
*Marcos Adán Ruiz Rodríguez*

Una de las principales actividades que realiza el personal encargado de la vigilancia epidemiológica es el estudio de brotes o epidemias de eventos adversos para la salud. Actualmente, la metodología para estudiar un brote se ha convertido en una herramienta básica para la práctica de la salud pública, ya que con ella es posible identificar los factores asociados con la presencia del evento adverso, implantar medidas de control para reducir el impacto del brote, obtener información básica para establecer la etiología del evento y transmitir información seria y oportuna a la población con el fin de establecer medidas de control o prevenir eventos futuros.

La investigación de brotes está formada por una categoría especial de estudios epidemiológicos especialmente orientados al examen de aquellos eventos de salud que tienen una presentación súbita, aguda, de ocurrencia inesperada y que por su naturaleza requieren de una respuesta inmediata. Como ejemplos de este tipo de problemas se encuentran los brotes de intoxicación alimentaria o los de enfermedades febriles exantemáticas.[1,2,3]

Un brote se define como la presencia de un número de eventos adversos para la salud mayor al esperado para un lugar y periodo determinados.[4] También se ha definido como una modalidad epidémica que se limita a "un aumento localizado en la incidencia de una enfermedad".[5]

En el ámbito de la salud pública, en especial en la vigilancia epidemiológica, la definición de brote y epidemia se consideran similares. Véase así que la epidemia se define como la aparición de un exceso en el número de casos de enfermedad, conductas u otros eventos relacionados con la salud en relación con los que normalmente se esperarían en una comunidad o región.[5] En ambos términos se reconoce la presencia de las tres variables básicas de la epidemiología observacional: tiempo, lugar y persona, es decir, las que permiten describir la magnitud, distribución y extensión del problema.

Frecuentemente se utiliza como sinónimo de brote el concepto de *cluster o cúmulo* de eventos en salud o casos pero, aun cuando este concepto se define como dos o más casos que se presentan juntos, ya sea en lugar, en tiempo o de manera combinada, preferentemente se refiere a eventos de baja frecuencia de carácter crónico, tales como cáncer o defectos al nacimiento. Brownson[6] describe algunas diferencias entre ambos conceptos (Cuadro I).[6,7,8]

El estudio de brote dentro de la epidemiología de la salud pública no es algo nuevo. Los estudios asociados a la comprensión de la peste o "muerte negra"—cuyo agente infeccioso es la *Yersinia pestis*— fueron documentados desde que la epidemia hiciera su primera aparición en Europa, alrededor de 1330. Entre otros, destacan los relatos contemporáneos de Giovanni Boccaccio, Geoffrey Chaucer, Gilles di Musis y de Agnola di Tura.[9]

**Cuadro I Comparación entre la investigación de brote y de *cluster*.**

| *Características* | *Investigación de brote* | *Investigación de cluster* |
|---|---|---|
| Enfermedad/condición | Infecciosa | No infecciosa |
| Agente etiológico | Transmisible | A menudo desconocido o agentes combinados |
| Periodo de tiempo para el desarrollo de la investigación | Corto (días o semanas) | Largo (semanas o meses) |
| Estimación del efecto | Generalmente de moderado a alto | Generalmente de bajo a moderado |
| Nivel de exposición | Moderado a alto | Bajo |
| Periodo de exposición | Agudo (horas o días) | Crónico (años o décadas) |
| Confirmación de laboratorio de casos y/o exposición | Común | Menos común |
| Posibilidad de establecer al relación causa/efecto | Moderada a elevada | Baja |
| Diseño de estudio más utilizado | Estudio de cohorte retrospectiva | Estudio de caso-control |

Fuente: Referencia 6



En México se tienen registros de la presencia de brotes de gran magnitud desde la conquista del imperio azteca por los españoles hacia 1521. Estos procesos devastaron a la población con enfermedades como la viruela (hueyzahuatl) en 1521, el sarampión (tepitonzahuatl) en 1538, o bien la rabia canina silvestre en 1709. A partir de la conquista de México, brotes de tifus (matlazahuatl) aparecen episódicamente y se reportan epidemias en 1526, 1533, 1536, 1564, 1588 y 1596. A lo largo de la historia de este país se han registrado numerosos brotes de distintos eventos en salud, pero destacan los padecimientos infecciosos.[10,11]

El estudio sistemático y cuantitativo de los brotes tiene su punto de partida en el clásico estudio del brote de cólera en Londres de 1854 realizado por John Snow. En esta investigación, con base en un trabajo epidemiológico fuera de oficinas o, como lo llamaron en alguna ocasión, "shoe-leather epidemiology", se caracterizan perfectamente los síntomas de la enfermedad, se describe a detalle la distribución de los casos del padecimiento en la vecindad del Soho; y de hecho, se señala la distribución de los mismos síntomas en las calles cercanas a las bombas de distribución de agua. Este estudio también desarrolla trabajo de campo en el que se identifican enfermos y se determina el origen del agua de consumo, lo cual permitió establecer la fuente de contaminación y proponer medidas de intervención para contener la epidemia.[12,13]

El trabajo realizado por Snow se ha perfeccionado con el transcurso de los años permitiendo la identificación de brotes y epidemias de gran importancia que han ocurrido tanto en el siglo XX como en el actual; como ejemplo de ello se tiene el brote de ántrax asociado con liberación intencional de *Bacillus an-*

*thracis* registrado en el año 2001, o el brote de síndrome respiratorio agudo reportado a la Organización Mundial de la Salud el 11 de febrero del 2003 y que hoy se conoce como síndrome respiratorio agudo severo (SARS por sus siglas en inglés), el cual ha tenido impacto tanto en la salud de las poblaciones como en la vida económica y social de los países afectados.[14,15]

Hoy en día, el avance en la tecnología de medios de transporte y comunicación tiene como consecuencia que el intercambio comercial y la movilidad poblacional entre países sean más eficientes sin importar las distancias geográficas; sin embargo, esto genera el incremento en las posibilidades de emergencia de problemas de salud nuevos o la re-emergencia de los que ya se habían erradicado.[16] El desarrollo de un brote —además del impacto en la salud de la población que naturalmente posee— puede traer consigo consecuencias importantes de otra índole en la vida de la región o país, dependiendo el tipo de problema.

Un ejemplo reciente ha sido la detección y estudio del brote de SARS en Canadá a un mes después de haberse detectado y reportado los primeros casos en China. El caso índice del brote en Toronto procedía de una familia china radicada en esta ciudad y que habían ido a visitar familiares a China durante el mes de febrero, tiempo en el cual el brote en ese país se encontraba activo.[17] A partir del reporte de estos casos, las repercusiones económicas no se dejaron esperar tanto por la disminución de los visitantes o cancelación de eventos con afluencia de visitantes de diversas partes del mundo.[18]

En este capítulo se discute el método básico para el estudio de brotes así como las generalidades para el control de los mismos y recomendaciones que posibiliten a los estudiantes o trabajadores de la salud pública contar con una guía que facilite el desempeño de su trabajo en esta área.

Frecuentemente los brotes son notificados por diferentes vías: la población afectada notifica a los servicios de salud; el personal de las unidades de salud donde ocurre el brote identifica qué tipo de brote es; los medios de comunicación —radio, televisión, prensa— lo denuncian, y los epidemiólogos trabajan el seguimiento de tendencias de los padecimientos que están incluidos en los sistemas de vigilancia epidemiológica con el fin de alertar sobre cualquier problema.

La investigación del brote es realizada regularmente por el personal responsable de los programas de atención de emergencias epidemiológicas y se reconocen en el trabajo directo de indagación tres tipos de actividades que agrupan al conjunto de acciones sobre el brote: a) la investigación formal de evento, b) la implantación de medidas de control y, c) el seguimiento del impacto de dichas acciones.

La investigación de un brote no es un asunto fortuito o que se desarrolla como una respuesta desordenada; por el contrario, en la mayoría de los servicios de salud se cuenta con una unidad de atención de brotes o emergencias epidemiológicas con una infraestructura mínima que permite dar una respuesta inmediata y ordenada al evento. Estas unidades de respuesta se encuentran preferentemente a cargo de personal especializado en las áreas de epidemiología, salud pública y comunicación de riesgo, los cuales, como punto de partida antes de salir a efectuar investigación de brote, determinan claramente quién será el responsable de conducir las acciones del grupo de trabajo en campo.

El trabajo de campo se planifica cuidadosamente, por lo que dentro de las acciones de preparación se debe considerar en detalle lo siguiente:

1. Contar con una invitación para participar en la investigación del evento. Esto es pertinente cuando se encuentran in-

volucradas responsabilidades estatales, regionales y/o jurisdiccionales de manera combinada.
2. Siempre es deseable, antes de salir a realizar la investigación, notificar a las autoridades internas del sitio donde ocurre el brote sobre la futura presencia del equipo de trabajo, para evitar conflictos comunitarios o de otra índole durante la operación en campo.
3. El siguiente paso es asegurar la independencia operacional en transporte, comunicaciones y alojamiento para el grupo destinado a la investigación.
4. Después se debe realizar la revisión exhaustiva de los implementos a utilizar en campo: desde la papelería necesaria para el trabajo hasta el equipo informático y de comunicación.

Dos puntos clave a fin de recibir el apoyo necesario en caso de así requerirse durante el proceso de indagación del evento son: tener identificadas y aseguradas las potenciales colaboraciones con expertos en diversas áreas (enfermedades infecciosas, enfermedades crónicas, evaluación de riesgo, entre otros); así como el vínculo con laboratorios donde solicitar el análisis de muestras de diversos modelos biológicos (sangre, orina, etc.) o de matrices ambientales (agua, polvo, lodo, alimentos) que faciliten la identificación de las potenciales fuentes de origen del brote.[19,20,21]

## LA INVESTIGACIÓN DEL BROTE EN EL CAMPO

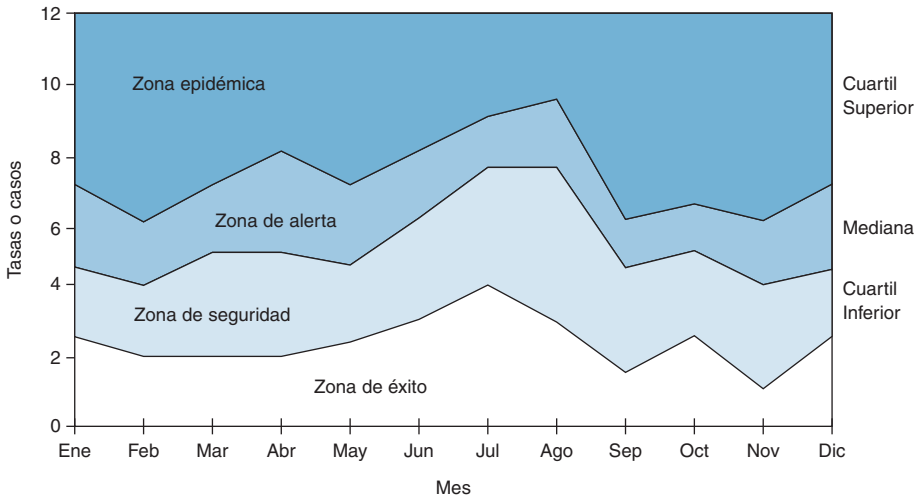


El equipo de trabajo que realiza la investigación del brote deberá cubrir de manera ineludible las siguientes líneas de acción:

1. Determinar la existencia del brote
2. Confirmar el diagnóstico
3. Definir quién será considerado un caso y estimar el número de casos
4. Orientar y analizar la información en tiempo, lugar y persona
5. Determinar quién está en riesgo de enfermar
6. Desarrollar una hipótesis explicatoria
7. Comparar la hipótesis con la información recolectada
8. Preparar un estudio sistemático
9. Preparar un reporte escrito
10. Ejecutar medidas de control y prevención.[22]

Estas líneas de acción de investigación del evento no se contraponen con la implantación y ejecución de medidas generales de control y prevención. Ambas actividades pueden realizarse de manera simultánea a fin de evitar un incremento en el número de casos del evento que se estudia. En algunas ocasiones el equipo de trabajo llega al sitio e instala actividades de control; algunas de ellas antes de probar la hipótesis que haya sido formulada, e.g. desalojo de la población ante un escape de químicos o la clausura de comercios expendedores de algún alimento sospechoso.

**1. Determinar la existencia del brote.** Con la información preliminar disponible sobre la presencia del brote que se recibe de las fuentes de información formal o no formal, se hace necesario revisar en primera instancia si realmente existe un exceso en el número de casos del evento. La herramienta útil para ello se encuentra en la información del canal endémico que proporciona el sistema de vigilancia epidemiológica específico del evento. Al tener la cifra preliminar de casos sospechosos del evento y el canal endémico, el equipo de trabajo




*Fuente*: Referencia 23

**Figura 1** **Ejemplo de canal endémico**

puede tener una vista rápida comparativa del número de casos reportados como brote con respecto a los esperados.

El canal endémico es la representación gráfica de la frecuencia esperada de casos de un evento determinado (figura 1) y se construye a partir del comportamiento semanal o mensual del evento durante los últimos cinco años. Como resultado de su construcción con base en el método de la mediana y los cuartiles es posible identificar cuatro zonas: éxito, seguridad, alarma y epidémica.[23] Con la frecuencia de casos reportada en el evento a investigar y contando con este canal endémico se puede establecer si esta frecuencia de casos se encuentra en la zona epidémica, esto es, si ha rebasado el número de casos esperados. En los servicios de salud se cuenta con los canales endémicos principalmente de las enfermedades de notificación inmediata o que son de importancia para la salud pública de la región o entidad.

Es conveniente puntualizar que un número excesivo no necesariamente se refiere a cifras muy grandes; alude a la frecuencia relativa con la que normalmente se presenta el evento en investigación. El impacto de la enfermedad en sí misma puede ser trascendente para el área, lo que obliga a establecer medidas inmediatas de investigación y control. La presencia de eventos relevantes en salud pública —ya por su impacto potencial en la salud poblacional o por haber estado ausentes durante mucho tiempo en la región—, así como la reaparición de brotes en forma de caso único, se considerarán brotes reales y deben ser estudiados como tales con las medidas de control inmediatas apropiadas. Un ejemplo sería la presencia de un caso sospechoso de poliomielitis en cualquier país, haya erradicado o no la enfermedad [24,25] o los casos de ántrax cutáneo registrados en los Estados Unidos inmediatamente después del atentado al World Trade Center de Nueva York. En este último

brote, después de haber detectado cuatro sobres que contenían esporas de *Bacillus anthracis* que fueron repartidos a través del Centro de Distribución Postal de Nueva Jersey hacia las ciudades de Nueva York y Washington, el equipo de campo confirmó la presencia de un caso de ántrax cutáneo e implantó las actividades de investigación e intervención del estudio de brote.[26]

Cuando eventos con este grado de importancia en salud pública se presentan en número de dos o más casos asociados epidemiológicamente en tiempo y lugar,[5] deben ser considerados como parte de un brote y recibir una atención inmediata y especial para controlarlo.

Es importante definir adecuadamente los límites del área geográfica de estudio, ya que el trabajo y tiempo dedicado a recoger la información es directamente proporcional al periodo de tiempo transcurrido y al tamaño de área geográfica definida para el estudio del brote. Frecuentemente la incidencia del padecimiento fluctúa con la estacionalidad, por lo que es necesario hacer las comparaciones incluyendo los datos de periodos similares de años previos; de aquí la utilidad de contar con una descripción detallada de canal endémico.

Cuando se trata de eventos que no sean de notificación obligatoria y no estén incorporados a los sistemas de vigilancia, las fuentes de información que pueden ser utilizadas en el área de estudio son las historias clínicas hospitalarias, la hoja diaria del médico en las unidades de salud dentro del área del brote, el registro de certificados de defunción, los testimonios de miembros de la comunidad, etc. Sin embargo, lo más importante es establecer si el número registrado de casos es superior al esperado o si aquéllos se han presentado en áreas diferentes a los sitios donde normalmente ocurren.

**2. Confirmar el diagnóstico.** El realizar el diagnóstico del evento que se indaga representa una de las etapas más importantes y debe basarse, en la medida de lo posible, en los signos y síntomas que de manera persistente se identifican desde el inicio del reporte del brote. En esta parte es importante el papel que pueden jugar especialistas relacionados con el evento generador del brote para recibir orientación sobre el cuadro clínico, variantes de éste, el mejor tratamiento y sobre las pruebas diagnósticas que deben llevarse a cabo para una confirmación diagnóstica. Si con los primeros casos se pueden realizar los análisis de laboratorio que complementen el diagnóstico, ya sea en muestras humanas o ambientales, dependiendo el tipo de evento, se podrá ayudar en la implantación oportuna de medidas de control.

En el área de trabajo es importante que el líder del grupo de trabajo y los epidemiólogos visiten y entrevisten de manera directa a algunos casos y líderes comunitarios, a fin de obtener información que contribuya a la investigación. El contacto con los enfermos dará al equipo una idea clara sobre la percepción comunitaria y dimensiones del problema en términos no sólo de salud, sino de impacto social.

***3.* Definir quién será considerado un caso y estimar el número de casos**. Regularmente, con la información que se ha recopilado durante el proceso de investigación del brote, el grupo de trabajo puede desarrollar una definición de caso del evento en salud en estudio. Esta definición es preliminar e inicialmente se construye con base en los siguientes tipos de variables: epidemiológicas (tiempo, lugar y persona) y clínicas (signos y síntomas que se describen por los casos). Las definiciones de caso construidas con estas variables no son suficientes para iniciar la intervención pero permiten orientar las acciones primarias de contención, por lo que es de mucha utilidad, en la medida de lo posible, que a la definición de caso se incorpore la variable "resultados de la-

boratorio" (p.e. nivel de anticuerpos del evento o aislamiento de agente causal).

Al disponer de esta información, se pueden construir por lo menos las siguientes definiciones de caso: caso sospechoso, caso probable y caso confirmado. Esta clasificación es también importante porque nos permitirá el manejo epidemiológico de las personas que aun estando involucradas en el brote puedan ser definidas como "no casos".

El caso sospechoso es aquél que cumple con alguno de los elementos de la definición clínica. El caso probable es todo aquél que sea compatible con la definición clínica del evento. El caso confirmado es aquél que cumple con todos los criterios de clasificación propuestos y que será reportado al sistema de vigilancia epidemiológica.

Para ejemplificar el uso de la definición de caso utilizaremos como evento a investigar el sarampión. La definición clínica de caso que se utiliza en la investigación de brotes y que también es considerada para la vigilancia epidemiológica es: una enfermedad caracterizada por un eritema generalizado de tres o más días de duración, fiebre de o mayor a 38.3ºC, tos, coriza o conjuntivitis. El caso sospechoso entonces, es toda aquella persona que, en la zona de estudio, reporte enfermedad febril acompañada de *rash*. El caso probable es aquella persona que sea compatible con la definición clínica. Y el caso confirmado es todo aquel caso en el que se confirma por laboratorio la presencia de anticuerpos contra el sarampión, y que es compatible con la definición clínica de sarampión o que esté relacionada epidemiológicamente con otro caso confirmado. Se encuentran disponibles las definiciones de caso para vigilancia epidemiológica y que son útiles para la investigación de brotes tanto para eventos infecciosos como para otro tipo de eventos.[27,28]

Se recomienda que la definición de caso sea sencilla y de fácil aplicación. Sin embargo, en estas situaciones en general se prefiere trabajar preferenciando la sensibilidad de la clasificación. Es importante recalcar que la definición operacional de caso que se utilice debe estar adecuada a las condiciones en las que se lleva a cabo la investigación del brote. La idea central de este procedimiento es desarrollar una serie de criterios que permitan definir si una persona tiene o no el evento en estudio, en el contexto particular de la situación en la que se lleva a cabo la investigación.

**4. Orientar y analizar la información en tiempo, lugar y persona.** La información compilada en el área del evento así como la que contiene la definición de caso permite que cada caso tenga información epidemiológica descriptiva, como fecha de inicio del evento, lugar de residencia y algunas características sociodemográficas que permitirán describirlos.

En el caso de la variable tiempo, la mejor descripción gráfica de la presencia del brote la ofrece una curva epidémica. La definición más simple de una curva epidémica es que ésta es la representación gráfica del número de casos durante la investigación del brote de acuerdo con la fecha de inicio de la enfermedad y con el número de casos documentados.

Esta representación gráfica es útil, ya que ofrece información del patrón de propagación del evento durante el brote, la magnitud del mismo, permite reconocer la presencia de casos aislados, observar si el evento tiene una tendencia en el tiempo y calcular el periodo de exposición del evento.

En la curva epidémica se pueden reconocer los siguientes elementos: a) periodo mínimo de exposición (la diferencia entre la fecha de aparición del caso primario y la duración máxima del periodo de incubación, b) periodo máximo de exposición (diferencia entre la fecha de aparición del último caso y el periodo mínimo de exposición), c) promedio del periodo de incubación (diferencia entre el día del pico máximo del brote y la media del periodo de incu-

bación), d) pico máximo del brote (fecha en la cual se tiene el mayor número de casos registrados), e) caso índice (el primer caso que se detecta durante la investigación del brote) y f) caso primario (es el individuo que introduce el evento en el grupo afectado adquiriendo la enfermedad de la fuente original, no necesariamente es la primera persona diagnosticada que aparece en el brote). (Figura 2)

El patrón de propagación del evento dentro de una curva epidémica permite que se clasifique al brote como de origen común, propagado o mixto. En el brote de origen o fuente común, la población está expuesta de manera intermitente o continua y los periodos de exposición serán cortos o largos. La curva epidémica de un brote de fuente común y exposición intermitente muestra picos irregulares que son reflejo del tiempo de exposición; un ejemplo de este tipo de brote es factible que se presente en un brote por intoxicación alimentaria reportado por Emery *et al*[29] (figura 3). En la curva epidémica de fuente común y exposición continua, los casos se incrementan gradualmente y más que picos es factible encontrar una meseta[30] (figura 4). Por último, la curva epidémica propagada es aquella que pasa de persona a persona y se presentan picos progresivamente más altos debido a que cada uno de estos picos tiene un periodo de incubación diferente[31] (figura 5 y 6).

Describir la distribución del brote en mapas ubicando gráficamente en ellos los casos —el domicilio de los afectados, el área interior



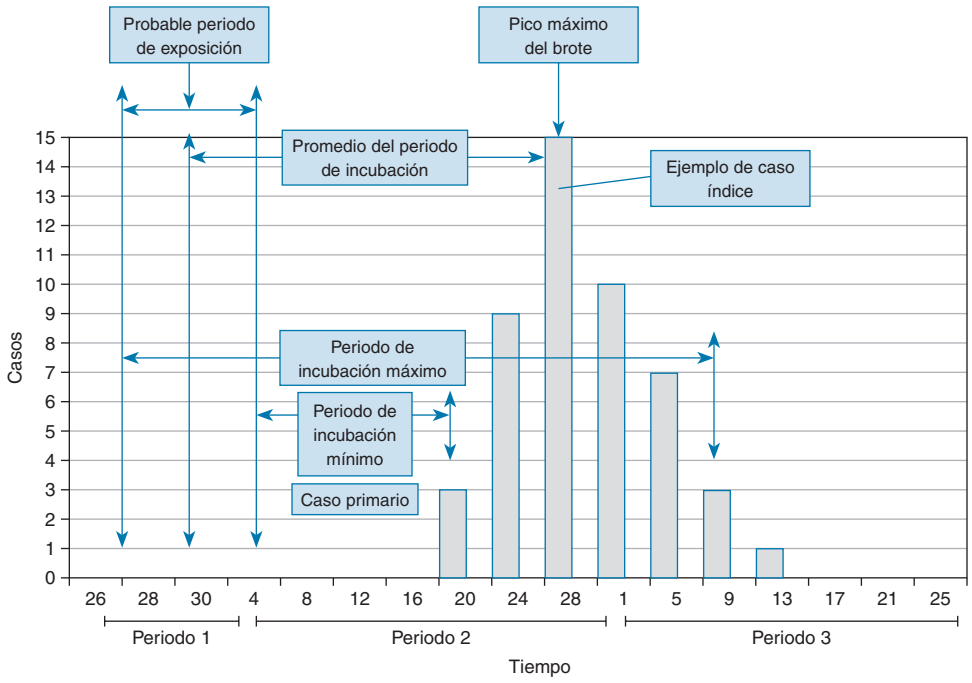
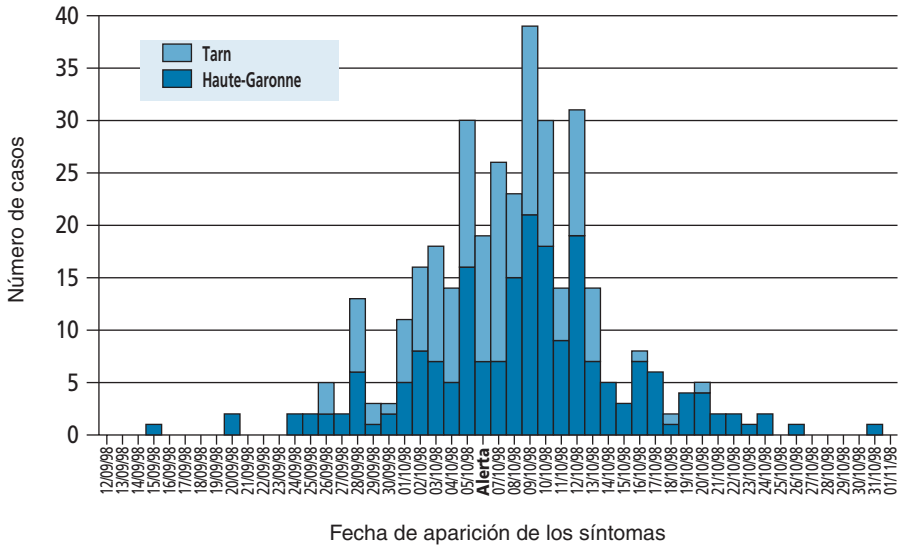



**Figura 2** **Características de la curva epidémica**




*Fuente*: Referencia 29

**Figura 3 Curva epidémica de un brote de fuente común con exposición intermitente**

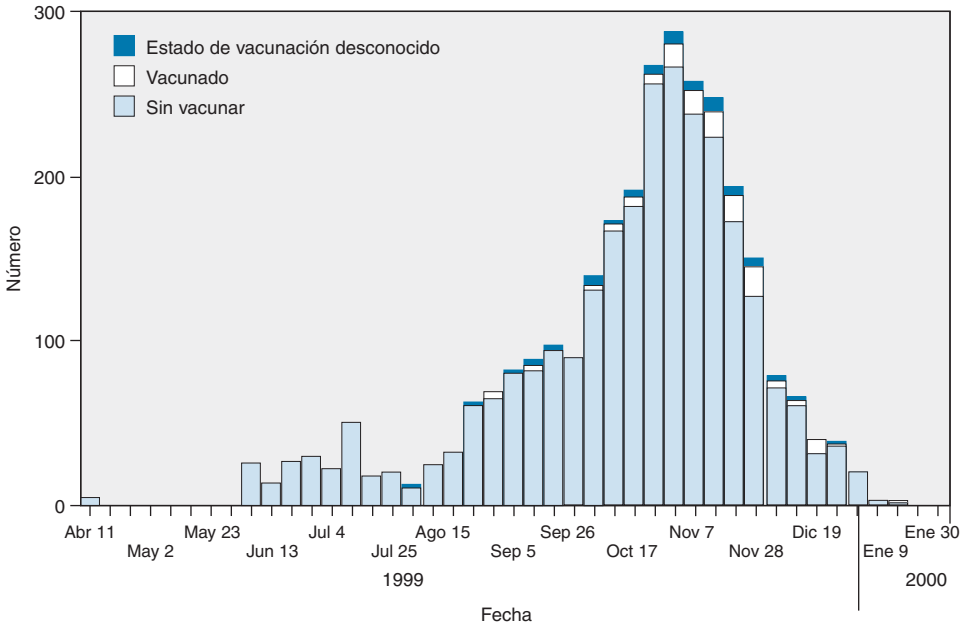


donde se registra, en caso de que el brote haya ocurrido dentro de una institución, por ejemplo en un hospital— permite evaluar la extensión geográfica del problema, y si a ello sumamos la información de las posibles causas (por ejemplo: fuentes de agua, alimentos, las líneas de abastecimiento de agua, gases medicinales, etc.), así como la fecha de inicio del evento, entonces se pueden trazar rutas de transmisión, ubicando las probables fuentes de origen del evento o reservorios. Con la información relativa a diversas características de los casos (como edad, sexo, nivel educativo, ocupación, etc.) podrá describirse a la población afectada.

**5. Determinar quien está en riesgo de enfermar.** Tanto el análisis de la curva epidémica, como el conocer el número de enfermos, cuándo y dónde enfermaron y las características clínicas del padecimiento, puede encaminar al investigador a identificar cómo y por qué se inició el brote. A partir de esta información se pueden implementar otras acciones de control del brote que pueden ir desde la toma de muestras de alimentos hasta la clausura de establecimientos.

**6. Desarrollar una hipótesis explicatoria.** Éste es el primer paso dentro de la epidemiología analítica del estudio de brotes que implica la evaluación metódica de las evidencias epidemiológicas, clínicas y de laboratorio, así como de las hipótesis que expliquen las exposiciones específicas que originaron el brote. En algunas ocasiones los resultados del estudio analítico encuentran que la exposición no está relacionada con la enfermedad y entonces se deberá plantear una nueva hipótesis.

En esta fase regularmente el brote se encuentra en su fase álgida y se requiere obtener

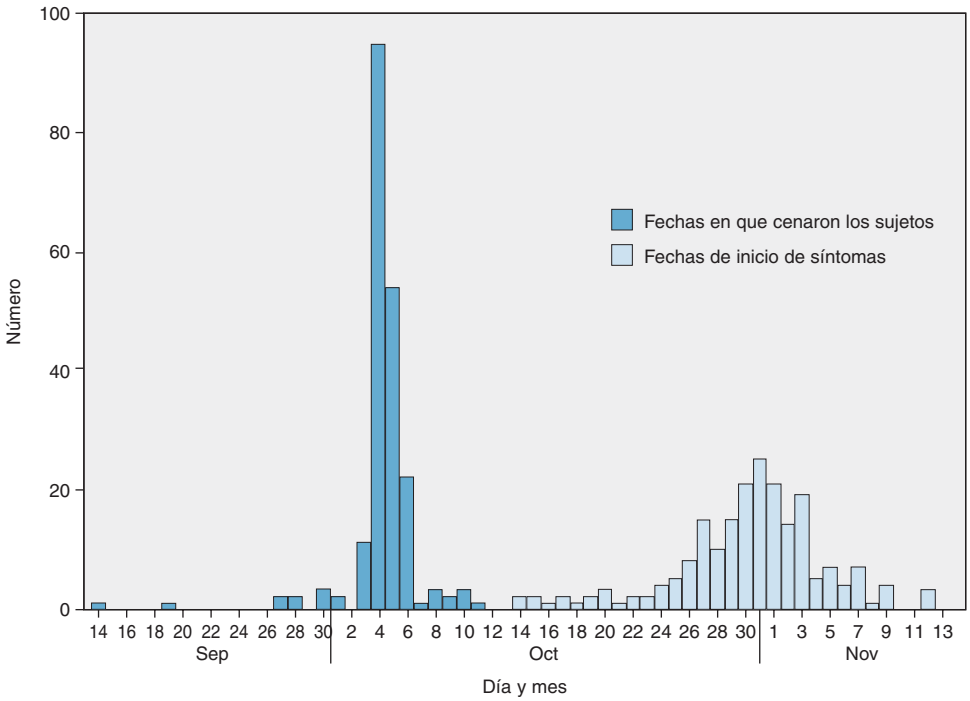



*Fuente*: Referencia 30.

**Figura 4** **Curva epidémica de un brote fuente común y exposición continua**

información que permita orientar las acciones de control, existiendo por otra parte presión importante de parte de los actores participantes del evento (población, autoridades de la región y medios de comunicación, entre otros), los cuales están requiriendo información sobre el evento y medidas para contenerlo. Con toda esta presión para obtener resultados y una hipótesis construida llega el momento de iniciar el trabajo en sitio. Este trabajo en sitio, o trabajo de campo, es la parte central para indagar la magnitud y los factores de riesgo que originaron el brote. Aun cuando no se cuente con las condiciones óptimas para el desarrollo de una investigación planeada, el estudio de brote debe cubrir todos elementos propios que propone la metodología de investigación en epidemiología.

En la investigación de un brote los diseños epidemiológicos que se utilizan con más frecuencia son el de cohorte retrospectiva y el de casos y controles, dado que el evento ya se presentó o está activo. En el diseño de cohorte retrospectiva se busca reconstruir la experiencia de exposición de los miembros de la cohorte y se mide la asociación por medio de la razón de riesgo o riesgo relativo comparando las tasas de ataque de las personas expuestas y no expuestas de la cohorte. El diseño de casos y controles se ha popularizado en la investigación de brotes por las ventajas que ofrece: e.g. los resultados se obtienen rápidamente, son de bajo costo y se pueden estudiar varias exposiciones de manera simultánea.[32] Las precisiones metodológicas para desarro-




*Fuente:* Referencia 31

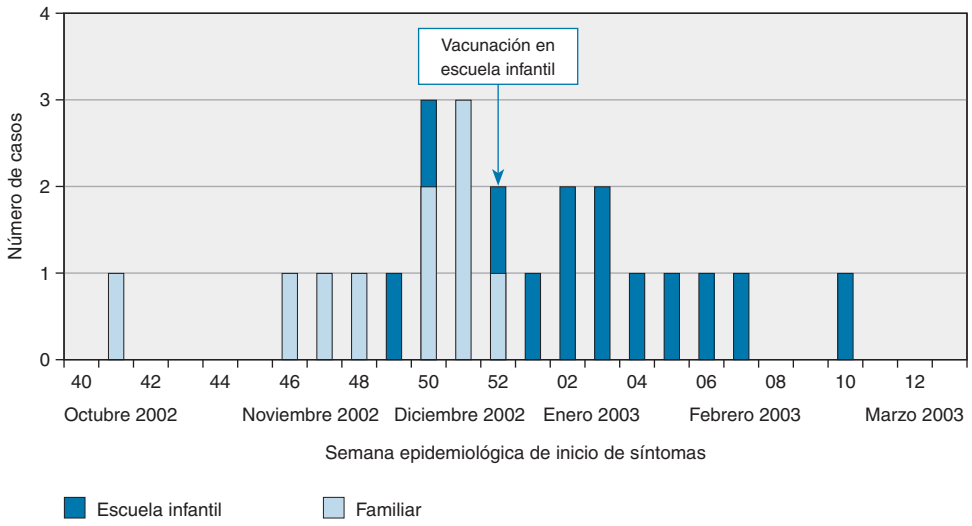
**Figura 5 Curva epidémica de un brote fuente propagada**

llar ambos tipos de diseño se han revisado en los capítulos 5 y 6.

**7. Comparar la hipótesis con la información recolectada.** Con la información obtenida el grupo de investigación realizará la comparación con la hipótesis propuesta y buscará establecer si los resultados obtenidos del evento (cuadro clínico, periodo de incubación, población afectada) coinciden con resultados ofrecidos por otros estudios. En caso contrario se revisará la hipótesis a la luz de los hallazgos.

**8. Planear un estudio más sistemático.** Después de concluir la fase de la investigación del brote, donde se cuenta con resultados que orientan al factor de riesgo generador del evento, la magnitud del mismo en términos de la población afectada, o bien de su letalidad, es posible que se requiera desarrollar un estudio epidemiológico más sistemático que cubra algunos elementos que, por la premura de tiempo en obtener resultados, se hayan pasado por alto en la investigación de campo. En esta fase pueden ser utilizados los diseños epidemiológicos comentados.

**9. Preparar un reporte escrito.** El hecho de documentar y publicar la investigación de un brote tiene diversas finalidades, entre las que resalta la toma de decisiones. Por otro lado, este reporte puede convertirse en un invalua-




*Fuente*: Referencia 33

**Figura 6 Curva epidémica propagada**



ble apoyo para prevenir brotes futuros o bien justificar la necesidad de un nuevo programa de intervención o reforzamiento del sistema de vigilancia epidemiológica.

En no pocas ocasiones se presentan controversias sobre posibles daños a la salud por diversos eventos. Un reporte bien escrito, claro y objetivo puede ayudar a resolver dichas polémicas y a la vez sería un valioso instrumento para aprender y enseñar epidemiología. De hecho, lo más recomendable es someter a publicación el reporte de brote en revistas especializadas o en boletines epidemiológicos de tal forma que se logre socializar el conocimiento obtenido en la investigación de este tipo de eventos.

**10. Implementar las medidas de prevención y control correspondientes.** El objetivo final de la investigación de un brote es la aplicación de medidas preventivas y de control que eviten o limiten los daños a la salud. Esta implantación estará orientada en función de los resultados de la investigación epidemiológica y de los resultados de los exámenes realizados en las muestras humanas y ambientales, pero en general se pueden reconocer tres tipos de medidas de intervención: eliminar la fuente que dio origen al brote, interrumpir la propagación de la fuente a los susceptibles y proteger a los susceptibles de las consecuencias de la exposición. En el brote de SARS de Toronto, que inicialmente fue intrahospitalario, los servicios de salud locales establecieron una serie de medidas entre el personal hospitalario y la comunidad así como con los casos confirmados, apoyadas por una intensa campaña de medios; estas medidas de control impactaron el trabajo hospitalario retrasando actividades programadas, al dar preferencia a las actividades de control de SARS, lo cual permitió la remisión del brote aproximadamente des-

pués de 120 días del registro del primer caso.[34] La implantación de medidas de control contribuye a la verificación de la hipótesis desarrollada. Por otro lado, actuar precipitadamente (por ejemplo, asegurar productos alimentarios, cerrar un restaurante, prohibir la importación de frutas y otros productos) puede causar graves daños a los actores económicos y sociales. Pero, aun con esta premisa, durante el estudio de un brote la responsabilidad central del equipo que realiza la investigación y la intervención es proteger la salud de la población.

La investigación de brotes está influida por el tipo de evento que se presenta, la rapidez de propagación y la oportunidad de intervención. En todo proceso de control de brotes el papel que juegan los medios de comunicación para coadyuvar a transmitir información adecuada y oportuna a la población es importante, de tal forma que la participación del epidemiólogo en la formulación de mensajes a la población adquiere una relevancia mayúscula.[35]

En el escenario sanitario de países con niveles económicos bajos, como es el caso de los países latinoamericanos, hay una mezcla de enfermedades infecciosas, crónicas, emergentes y reemergentes; y en todas ellas es potencialmente factible el desarrollo de brotes. Por esta razón, aun cuando la investigación de brotes ha sido considerada una actividad científica menor, es claro que posee una gran relevancia sanitaria y que se requiere fortalecer las habilidades del personal de salud para enfrentar estas contingencias. Por eso es necesario fortalecer las capacidades de los epidemiólogos en la investigación de brotes y con ello facilitar la intervención oportuna, el estudio dirigido y el control adecuado de estos importantes problemas de salud.[36]



## REFERENCIAS


1. Parrilla-Cerrillo C, Vázquez-Castellano J, Saldate-Castañeda O, Nava-Fernández L. Brotes de toxiinfecciones alimentarias de origen microbiano y parasitario. Salud Publica Mex 1993; 35(5):456-463.
2. Casanova-Cardiel L, Hermida-Escobedo C. Sarampión en el adulto joven. Hallazgos clínicos en 201 casos. RIC 1994; 46(2):93-98.
3. Sánchez BY, Álvarez PA, Saldaña E, Salomón A, Ruiz-Gómez J, Muñoz HO. ¿Qué pasa con la vacuna de sarampión en México? Estudio de un brote epidémico de sarampión. Bol Med Hosp Infant 1977; 34(2):291-296.
4. USDHHS, Investigating an outbreak. In: Principles of epidemiology: an introduction to applied epidemiology and biostatistics. Atlanta, GA: US Department of Health and Human Services, US Public Health Service, Centers for Disease Control and prevention, EPO, PHPPO, 1993: 435-525.
5. Last JM, ed. Dictionary of epidemiology (3rd Ed.). New York: Oxford University Press, 1995.
6. Brownson CR.Outbreak and cluster investigations. In: Applied epidemiology. Theory to practice. New York: Ross C. Browson and Diane Petitti, OUP, 1998.
7. Heath CW. Investigating causation in cancer clusters. Radiat Environ Biophys 1996;35: 133-136.
8. USDHHS, Public Health Services, Centers for disease control. Guidelines for investigating clusters of health events. Atlanta, GA: USDHHS, 1990;39(RR-11).
9. Stolley P, Lansky T. Investigating disease patterns: the science of epidemiology. Scientific American Library. 1995.
10. Somolinos d'Ardois G. Las epidemias en México durante el siglo XVI. Salud Publica Mex 1988;30(4):639-644.
11. Carrada T. Investigación documental de la pri-

mera epidemia de rabia registrada en la República Mexicana en 1709. Salud Publica Mex 1978; XX (6):705-716.
12. Annotations. A Snow centenary. The Lancet 1955; 268: 1113-1114.
13. Summers J. Soho - A history of London's most colourful neighborhood, Bloomsbury, London, 1989: 113-117. Tomado de: http://www.ph.ucla.edu/epi/snow/broadstreetpump.html.
14. MMWR. Update: Investigation of Anthrax associated with intentional exposure and interim public health guidelines, october 2001 october 19, 2001/50(41);889-893.
15. MMWR. Outbreak of severe acute respiratory syndrome. Worldwide, 2003. March 21, 2003/52(11);226-228.
16. Satcher D. Emerging infections: getting ahead of the curve. EID 1995; 1(1):1-6.
17. Poutanen S, Low D, Henry B, Finkelstein S, Rose D, Greene K, Tellier R *et al.* Identification of sever acute respiratory syndrome in Canada. N Engl J Med 2003;348:1995-2005.
18. Spurgeon D. Canada insists that it is a safe place to visit. BMJ 2003 May 3;326(7396):948.
19. World Health Organization. Epidemic and pandemic alert response, 2006. http://www.who.int/csr/alertresponse/logistics/en/index.html.
20. Gregg M. Conducting a field investigation. In: Field epidemiology. Oxford University Press, New York 1991.
21. Gregg M. The principles of an epidemic field investigation. In: Holland W, Detels R, Knox G. eds. Oxford textbook of public health. Vol.3 Investigative methods in public health. OUP 1985:284.
22. Goodman RA, Buehler JW, Soplan JP. The epidemiologic field investigation: science and judgment in public health practice. Am J Epidemiol 1990; 132(1):9-16.
23. Bortman M. Elaboración de corredores o canales endémicos mediante planillas de cálculo. Rev Pan Salud Pública 1999;5(1):1-8.
24. MMWR. Poliovirus infection in four unvaccinated children. Minnesota, august-october, 2005 october 14, 2005/54(Dispatch);1-3.
25. MMWR. Recommendations and reports. Poliomielitis prevention in the United States. Updated recommendations of the Advisory Committee on Immunization Practices (ACIP). May 19, 2000/49(RR05);1-22.
26. Greene C, Reeflhuis J, Tan Ch, Fiore A *et al.* Epidemiologic investigations of bioterrorism-related Ántrax, New Jersey, 2001. EID 2002: 8(10); 1048-1055.
27. MMWR. Case definitions for infectious conditions under public health surveillance.1997; 46(RR10).
28. MMWR. Case definitions for chemical poisoning. Jan 14, 2005;54(RR1).
29. Emery C, Haeghebaert S. New outbreak of trichinellosis in the Midi-Pyrénées region of France. September-october 1998. Eurosurveillance 1999; 4(1):13.
30. MMWR. Measles outbreak — Netherlands, april 1999–january 2000. April 14, 2000 (49): 14.
31. MMWR Hepatitis A outbreak associated with green onions at a restaurant — Monaca, Pennsylvania, 2003 november 28, 2003;52(47);1155-1157.
32. Dwyer MD, Strickler H, Goodman R, Armenian KH. Use of case control studies in outbreak investigation. Epidemiologic Reviews 1994;16(1):109-123.
33. Arce AA, Rodero GI, Iñigo MJ, Burgoa AM, Guevara AE. Brote de hepatitis A en una escuela infantil y transmisión intrafamiliar de la infección. Anales de pediatría 2004;60(3):222-227.
34. Svoboda T, Henry B, Shulman L, Kennedy E, Rea E, Ng W, *et al.* Public Health measures to control the spread of the severe acute respiratory syndrome during the outbreak in Toronto. N Engl J Med 2004;350:2352-2361.
35. Davidson JL, George LE. Nurses' use of the media to provide public health information during a hepatitis A outbreak. J of Professional Nursing 2004;(20)2:134-136.
36. Reingold A. Outbreak investigations: A perspective. EID 1998;4(1):21-27.

# Investigación de brotes — Ejercicios

*Óscar Velásquez**
*Mauricio Hernández Ávila***

## BROTE DE ENFERMEDAD EN UN ÁREA RURAL*

El martes 26 de agosto de 1986, se notificó a la Dirección General de Epidemiología de la Secretaría de Salud de México un brote de hepatitis en Huitzililla, una pequeña comunidad rural ubicada en el estado de Morelos. Por la información inicial se supo que, entre el 1 de junio y el 26 de agosto, se habían atendido aproximadamente 31 casos de enfermedad ictérica en el servicio de salud local. Los casos notificados se caracterizaron por un inicio abrupto de fiebre, anorexia, astenia, dolor abdominal y cefalea, seguidos de ictericia y coluria. El grupo de edad más afectado era el de 25 a 44 años, con 23 casos, sin que se hubiesen observado diferencias por sexo.

Se solicitó apoyo de investigación, por lo que en los primeros días de septiembre un grupo de cuatro médicos especialistas en epidemiología aplicada se trasladó a la localidad para reunirse con las autoridades de salud locales e iniciar la investigación del brote y aplicar medidas de contención.

1. ¿Qué es una epidemia y cómo se define operacionalmente?
2. ¿Puede usted determinar si en este caso se trata de una epidemia? ¿Qué información adicional requiere usted?
3. Entre los datos epidemiológicos que usted tiene ahora, ¿cuáles son poco usuales?
4. Formule un plan de acción para realizar el estudio correspondiente y tomar las medidas que se requieran.

## ANTECEDENTES

Huitzililla tiene una población de 1 757 habitantes, distribuidos en 20 manzanas. Recorren la comunidad tres pequeños arroyos que nacen de las filtraciones del Gran Canal de Tenango, procedente de Agua Hedionda. Uno de los arroyos es El Chapala, que corre de norte a sur; permanentemente tiene agua y sirve para el riego de terrenos en la comunidad; asimismo, se emplea para fines domésticos, es decir, para el lavado de ropa y trastes, para el aseo personal y como receptor de aguas negras. Otro es El Salto, que corre de oriente a poniente; tiene menos agua que el anterior, pero al llegar a los límites de la comunidad se mezcla con las aguas de las granjas establecidas, con lo que aumenta su caudal. Ambos arroyos desembocan en la Barranca de la Cueva. El otro arroyo nace en el noreste de la comunidad y se le conoce como El Sabino; el agua de este arroyo también es usada para fines domésticos y como surtidor de los pozos familiares. La gran mayoría de las casas cuenta con pozos propios para autoconsumo, los cuales son de escasa

* Secretaría de Salud, México.
** Instituto Nacional de Salud Pública, México.

profundidad debido a que los mantos freáticos se localizan cerca de la superficie. El 95% de las personas practica el fecalismo a ras del suelo, y las pocas letrinas existentes no fueron diseñadas para evitar la contaminación freática.

La población de la localidad es atendida por un médico pasante de servicio social, quien hace visitas una a dos veces por semana; asimismo, la comunidad cuenta con una promotora de salud que da atención comunitaria una vez por semana.

5. Enumere las posibilidades diagnósticas que deberán tenerse en cuenta.
6. ¿Cómo definiría usted un caso? Mencione una definición operacional.
7. ¿Considera que son éstos todos los casos?
8. ¿Cómo buscaría usted más casos y qué preguntaría?

## METODOLOGÍA PARA LA BÚSQUEDA DE CASOS

El 12 de septiembre, cuando los investigadores llegaron a la localidad, ya eran 74 los casos de enfermedad ictérica, todos ellos documentados y atendidos por el médico pasante. El equipo de médicos investigadores decidió realizar un censo poblacional de la localidad y, a partir del mismo, una búsqueda intensiva de casos de enfermedad ictérica con la siguiente definición operacional de caso: "Toda aquella persona que radicara en ese momento en Huitzililla y hubiera presentado ictericia a partir del 1 de junio y hasta el 29 de octubre de 1986." Con este procedimiento se identificó un total de 88 casos.



Adicionalmente, se estableció un sistema de vigilancia activa, de tal modo que el médico y la enfermera recorrieran la localidad una vez por semana buscando activamente casos nuevos de enfermedad ictérica. Esta estrategia permitió identificar 19 casos adicionales que no se incluyeron en el estudio intensivo del brote ya que fueron diagnosticados durante noviembre y diciembre, dos meses después de que se detectó el último caso directamente relacionado con la fase aguda del brote y de que se establecieron las medidas de control.

9. Usando los datos del cuadro I haga los gráficos que representen la distribución de los casos de acuerdo con el día y la semana en que fueron diagnosticados. Una vez hecho esto, caracterice el brote en tiempo, diga qué le sugieren estas gráficas, explique la presentación de casos que no son de fuente común y aclare qué información adicional necesitaría para determinar si se trata de una epidemia.

En el transcurso de los 12 meses anteriores al 1 de junio de 1986, se diagnosticaron en la comunidad de estudio cuatro casos de enfermedad ictérica; dos ocurrieron en una misma familia durante el mes de marzo de 1986, mientras que los restantes no tenían relación en tiempo y persona. La evidencia anecdótica de años anteriores también indicaba que la frecuencia anual de enfermedad ictérica en la comunidad oscilaba entre los 2 y los 5 casos por año.

En la figura 1 se presentan los casos de hepatitis notificados durante el año previo y los que corresponden al periodo actual; esta figura indica claramente la existencia de un brote epidémico.

Se concluyó que se estaba presentando un brote epidémico, puesto que el número de casos de ictericia era superior al número de casos habituales o esperados, tomando como referencia los casos del año anterior y la información anecdótica que se obtuvo en la comunidad. Asimismo, se

**Cuadro I** **Casos de enfermedad ictérica. Huitzililla, Morelos, junio-octubre de 1986**

| *Número de caso* | *Edad* | *Sexo* | *Manzana* | *Fecha de diagnóstico* |
|---|---|---|---|---|
| 1 | 13 | M | 1 | 05-06-86 |
| 2 | 3 | F | 1 | 15-06-86 |
| 3 | 10 | M | 5 | 15-06-86 |
| 4 | 34 | F | 5 | 15-06-86 |
| 5 | 22 | M | 6 | 18-06-86 |
| 6 | 56 | F | 5 | 20-06-86 |
| 7 | 49 | M | 6 | 20-06-86 |
| 8 | 9 | F | 5 | 28-06-86 |
| 9 | 30 | M | 10 | 30-06-86 |
| 10 | 15 | F | 5 | 01-07-86 |
| 11 | 31 | F | 5 | 05-07-86 |
| 12 | 25 | M | 9 | 05-07-86 |
| 13 | 18 | F | 12 | 07-07-86 |
| 14 | 18 | F | 12 | 07-07-86 |
| 15 | 43 | F | 13 | 08-07-86 |
| 16 | 53 | M | 18 | 11-07-86 |
| 17 | 36 | M | 4 | 15-07-86 |
| 18 | 60 | M | 8 | 19-07-86 |
| 19 | 18 | F | 12 | 20-07-86 |
| 20 | 34 | M | 4 | 20-07-86 |
| 21 | 34 | M | 5 | 20-07-86 |
| 22 | 22 | M | 5 | 20-07-86 |
| 23 | 19 | F | 15 | 22-07-86 |
| 24 | 21 | M | 1 | 25-07-86 |
| 25 | 31 | M | 15 | 25-07-86 |
| 26 | 35 | M | 10 | 28-07-86 |
| 27 | 28 | F | 1 | 30-07-86 |
| 28 | 20 | M | 5 | 30-07-86 |
| 29 | 18 | M | 1 | 31-07-86 |
| 30 | 20 | F | 12 | 31-07-86 |
| 31 | 26 | F | 18 | 01-08-86 |
| 32 | 29 | M | 4 | 01-08-86 |
| 33 | 4 | F | 5 | 02-08-86 |
| 34 | 17 | M | 3 | 02-08-86 |
| 35 | 28 | M | 9 | 02-08-86 |
| 36 | 18 | F | 6 | 03-08-86 |
| 37 | 57 | F | 1 | 0508-86 |
| 38 | 62 | M | 5 | 06-08-86 |
| 39 | 25 | M | 8 | 07-08-86 |
| 40 | 75 | F | 11 | 07-08-86 |
| 41 | 25 | M | 5 | 08-08-86 |
| 42 | 22 | F | 11 | 08-08-86 |

**Cuadro I** ***Continuación***

| *Número de caso* | *Edad* | *Sexo* | *Manzana* | *Fecha de diagnóstico* |
|---|---|---|---|---|
| 43 | 16 | M | 1 | 07-08-86 |
| 44 | 25 | M | 9 | 10-08-86 |
| 45 | 65 | F | 7 | 08-08-86 |
| 46 | 21 | F | 15 | 14-08-86 |
| 47 | 44 | M | 4 | 15-08-86 |
| 48 | 48 | F | 5 | 15-08-86 |
| 49 | 55 | F | 6 | 18-08-86 |
| 50 | 24 | F | 5 | 18-08-86 |
| 51 | 14 | M | 9 | 18-08-86 |
| 52 | 46 | F | 19 | 19-08-86 |
| 53 | 40 | M | 6 | 20-08-86 |
| 54 | 26 | M | 11 | 22-08-86 |
| 55 | 25 | F | 5 | 23-08-86 |
| 56 | 7 | M | 15 | 24-08-86 |
| 57 | 18 | F | 1 | 25-08-86 |
| 58 | 52 | M | 11 | 25-08-86 |
| 59 | 10 | F | 2 | 25-08-86 |
| 60 | 21 | F | 11 | 25-08-86 |
| 61 | 30 | M | 4 | 26-08-86 |
| 62 | 31 | F | 10 | 26-08-86 |
| 63 | 44 | M | 1 | 28-08-86 |
| 64 | 16 | F | 15 | 28-08-86 |
| 65 | 54 | F | 15 | 28-08-86 |
| 66 | 18 | F | 15 | 31-08-86 |
| 67 | 16 | M | 12 | 31-08-86 |
| 68 | 25 | F | 8 | 31-08-86 |
| 69 | 17 | F | 15 | 31-08-86 |
| 70 | 23 | M | 12 | 02-09-86 |
| 71 | 30 | M | 1 | 03-09-86 |
| 72 | 30 | F | 15 | 06-09-86 |
| 73 | 19 | M | 17 | 11-09-86 |
| 74 | 11 | M | 15 | 11-09-86 |
| 75 | 15 | F | 5 | 17-09-86 |
| 76 | 17 | M | 19 | 17-09-86 |
| 77 | 28 | F | 16 | 22-09-86 |
| 78 | 85 | F | 11 | 23-09-86 |
| 79 | 3 | F | 5 | 26-09-86 |
| 80 | 14 | F | 1 | 26-09-86 |
| *80 | 25 | F | 15 | 28-09-86 |
| 82 | 10 | M | 5 | 30-09-86 |
| 83 | 18 | M | 12 | 02-09-86 |

* Defunción

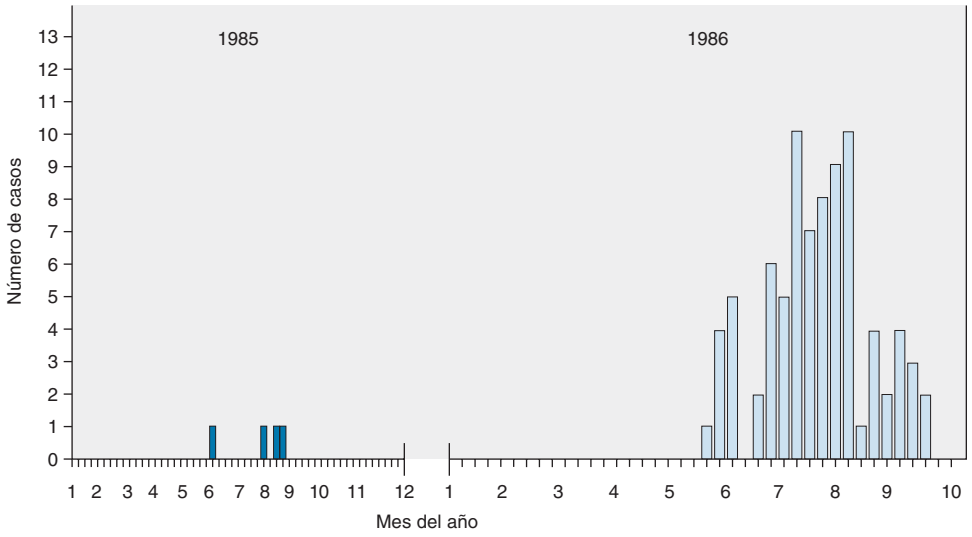



**Figura 1** **Distribución de los casos de hepatitis notificados entre 1985 y 1986 en Huitzililla, Morelos**



evidenció la distribución poco usual de los casos, en tanto que la mayoría de ellos correspondió a población adulta. En México, los brotes de hepatitis A ocurren por lo general en niños menores de 10 años, y los casos de hepatitis B, aunque se presentan con mayor frecuencia en adultos, raramente ocurren en proporciones epidémicas dentro de comunidades rurales con las características de Huitzililla.

Con la finalidad de profundizar en las características de la enfermedad, se diseñó un cuestionario individual donde se captaron, además de los datos demográficos, las características clínicas y la fecha de inicio, así como los antecedentes epidemiológicos de importancia para describir el brote.

**Cuadro II** **Casos de enfermedad ictérica y población en riesgo de acuerdo con la edad. Huitzililla, Morelos, 1986**

| *Grupo de edad* | *Número de casos* | *Población en riesgo* |
|---|---|---|
| Menos de 1 año | 0 | 62 |
| 1-4 | 3 | 235 |
| 5-14 | 9 | 583 |
| 15-24 | 32 | 333 |
| 25-44 | 30 | 348 |
| 45 y más | 14 | 196 |

**Cuadro III** **Casos de enfermedad ictérica y población en riesgo de acuerdo con el sexo. Huitzililla, Morelos, 1986**

| *Sexo* | *Número de casos* | *Población en riesgo* |
|---|---|---|
| Masculino | 44 | 900 |
| Femenino | 44 | 857 |

10. ¿Cree usted necesario realizar pruebas en el agua de los arroyos o pozos de la comunidad? ¿Qué indicador utilizaría?
11. Con la información actual, ¿podría usted concluir que el evento se trata de una epidemia de hepatitis?
12. Usando los datos de los cuadros II a V, estime la tasa de ataque por edad, sexo, manzana de residencia y localización de residencia en relación con los arroyos.
13. Utilice la información que tiene hasta el momento para proponer hipótesis tentativas en relación con la fuente de infección y el modo de transmisión.

**Cuadro IV** **Casos de enfermedad ictérica y población en riesgo de acuerdo con la manzana de residencia. Huitzililla, Morelos, 1986**

| *Manzana* | *Número de casos* | *Número de habitantes* |
|---|---|---|
| *1 | 2 | 62 |
| 2 | 1 | 46 |
| 3 | 1 | 76 |
| 4 | 5 | 110 |
| *5 | 19 | 113 |
| *6 | 5 | 102 |
| 7 | 1 | 18 |
| 8 | 3 | 50 |
| 9 | 4 | 52 |
| 10 | 3 | 75 |
| 11 | 6 | 173 |
| *12 | 7 | 31 |
| 13 | 1 | 81 |
| 14 | 0 | 70 |
| *15 | 11 | 228 |
| 16 | 1 | 160 |
| 17 | 1 | 75 |
| 18 | 3 | 86 |
| 19 | 4 | 111 |
| 20 | 0 | 47 |
| Total | 88 | 1757 |

* Manzanas cercanas a los arroyos

**Cuadro V** **Casos de enfermedad ictérica y población en riesgo de acuerdo con la localización de la residencia en relación con los arroyos. Huitzililla, Morelos, 1986**

| *Localización* | *Número de casos* | *Población en riesgo* |
|---|---|---|
| Cerca | 54 | 536 |
| Lejos | 34 | 1221 |

Las características en cuanto a la configuración y la duración de la curva epidémica y otros datos epidemiológicos sugieren un brote de hepatitis viral, en una primera etapa, de fuente común y continua y, posteriormente, por la transmisión de persona a persona. El grupo de investigadores consideró que la exposición para la primera etapa de esta enfermedad fue, aproximadamente, entre finales de mayo y principios de junio, cuando iniciaron las lluvias en la localidad.

El cuadro clínico fue muy característico de una hepatitis viral. Al analizar la información por grupo de edad se encontró que el más afectado fue el de 15 a 24 años; esta distribución sugiere que se trata de hepatitis no A no B, de tipo entérica, actualmente conocida como hepatitis tipo E. De acuerdo con la localización geográfica dentro de la comunidad, los casos se distribuyeron en 17 de las 20 manzanas que conforman la localidad. Se observó una marcada concentración de casos alrededor de dos pequeños arroyos que bordean la localidad, El Chapala y El Sabino, por lo que los investigadores consideraron que una posible fuente de exposición era precisamente el agua de los arroyos.



14. ¿Qué información adicional requeriría?
15. ¿Podría usted dar elementos para realizar un estudio de casos y controles a fin de confirmar algunas hipótesis?
16. ¿Qué información adicional requeriría sobre los pozos de agua?

El desarrollo de esta etapa constituyó la base para el diseño de un estudio de casos y controles, con los objetivos de identificar factores de riesgo asociados a la enfermedad y establecer el posible agente etiológico. El estudio se apoyó en exámenes de laboratorio que se realizaron en muestras de sueros sanguíneos para identificar: inmunoglobulina G (lgG) e inmunoglobulina M (lgM) para hepatitis A, así como antígeno de superficie (HBsAg), inmunoglobulinas totales anti-core (anti-HBc) e IgM anti-HBc, en presencia de anti-HBc positivo para establecer el diagnóstico de hepatitis B. Se obtuvieron muestras de heces de casos agudos para identificar virus por microscopía electrónica.

## ESTUDIO DE CASOS Y CONTROLES

Para obtener información más específica sobre las posibles hipótesis planteadas, se realizó un estudio de casos y controles con 32 casos primarios y con un grupo de 20 controles seleccionados aleatoriamente a partir del censo, con base en las variables de sexo, grupo de edad y residen-

**Cuadro VI** **Distribución de los casos y controles de acuerdo con las variables seleccionadas. Residentes de Huitzililla, Morelos, 1986**

| *Variables* | *Casos* | *Controles* | *Total* | *RM* (IC95%)* | *p* |
|---|---|---|---|---|---|
| Hierven el agua para beber | | | | | |
| Sí | 7 | 10 | 17 | 0.28 (0.085-0.92) | 0.0354 |
| No | 25 | 10 | 35 | | |
| Total | 32 | 20 | 52 | | |
| Antecedentes de contacto reciente con un caso de hepatitis | | | | | |
| Sí | 21 | 4 | 25 | 7.6 (2.1-27.1) | 0.0014 |
| No | 11 | 16 | 27 | | |
| Total | 32 | 20 | 52 | | |
| El pozo de donde obtienen el agua se encuentra al ras del suelo | | | | | |
| Sí | 12 | 6 | 18 | 1.4 (0.43-4.5) | 0.5802 |
| No | 20 | 14 | 34 | | |
| Total | 32 | 20 | 52 | | |
| El pozo de donde obtienen el agua se encuentra cubierto | | | | | |
| Sí | 12 | 11 | 23 | 0.49 (0.16-1.50) | 0.2164 |
| No | 20 | 9 | 29 | | |
| Total | 32 | 20 | 52 | | |
| El agua del pozo de donde obtienen el agua se ve contaminada a simple vista | | | | | |
| Sí | 29 | 4 | 33 | 38.6 (8.0-185.0) | 0.001 |
| No | 3 | 16 | 19 | | |
| Total | 32 | 20 | 52 | | |
| Se da algún tratamiento al agua del pozo antes de beberla | | | | | |
| Sí | 7 | 13 | 20 | 0.15 (0.04-0.51) | 0.0019 |
| No | 25 | 7 | 32 | | |
| Total | 32 | 20 | 52 | | |
| Antecedentes de contacto reciente con un caso de hepatitis | | | | | |
| Sí | 20 | 3 | 23 | 0.11 (0.03-0.42) | 0.0008 |
| No | 12 | 17 | 29 | | |
| Total | 32 | 20 | 52 | | |

*RM = Razón de momios

cia. Los resultados referentes a algunas de las variables consideradas en este estudio se presentan en el cuadro VI.

17. Resuma los datos sobre la encuesta de casos y controles del cuadro VI. Haga los cálculos para estimar la razón de momios, los intervalos de confianza y el valor de *p* para cada variable.

Se hicieron los mismos cálculos para todas las variables incluidas en el cuestionario y se descartaron algunas identificadas inicialmente como posibles factores de riesgo; por ejemplo, el consumo de ciertos alimentos en reuniones o eventos comunes y la aplicación de inyecciones, entre otras. El análisis de la información obtenida por la encuesta de casos y controles indicó la importante participación del agua como un vehículo de transmisión. El antecedente de hervir el agua disminuye significativamente el riesgo de enfermar. Los sujetos que notificaron hervir el agua presentaron una disminución en el riesgo equivalente a 3.5 veces, en comparación con aquellos que no lo hacen. De manera similar, los sujetos que señalaron tratar el agua antes de beberla (hirviéndola o clorándola) presentaron un riesgo cerca de nueve veces menor.

En contraste, el antecedente de haber detectado contaminación visible en el pozo familiar se asoció con un incremento de cerca de 40 veces en el riesgo de presentar el evento. También el antecedente de haber estado en contacto con un caso de hepatitis aumenta el riesgo de padecer esta enfermedad hasta siete veces.

Los datos anteriores sugieren que el mecanismo de transmisión probablemente incluyó la contaminación del agua, así como el contacto directo con los casos.



18. ¿Qué información adicional necesita para confirmar el diagnóstico, ahora que ya conoce el mecanismo de transmisión?

## ESTUDIOS DE LABORATORIO

En los estudios serológicos practicados en 62 muestras analizadas se observó que 97% de los pacientes presentaba IgG contra el virus de la hepatitis A, y en ningún caso se identificó IgM (cuadro VII). En cinco pacientes (9.4%) se detectaron anticuerpos IgG anti-HBc; paralelamente, fueron negativos para IgM anti-HBc. Tampoco hubo positivos para HBsAg.

19. ¿Cómo interpretaría usted los resultados del cuado VII? ¿Qué información adicional requeriría?

Mediante microscopía electrónica, hecha en el Laboratorio de Virología de los Centros para el Control y la Prevención de Enfermedades (CDC, por sus siglas en inglés), de Atlanta, Georgia, se identificaron partículas virales en las heces de dos pacientes con características morfológicas semejantes a las observadas en un brote de hepatitis no A no B de transmisión entérica estudiado en la misma época en Nepal (el tamaño aproximado de las partículas virales fue de 32-34 mm).

La combinación de cuatro sueros de pacientes con cuadro agudo de este brote se retó contra las partículas virales observadas en heces de enfermos por un brote que se presentó en la Unión Soviética; en ellas se observó una aglutinación de las mismas, empleando la técnica de inmunofluorescencia.

**Cuadro VII Resultados serológicos en 62 casos de enfermedad ictérica. Huitzililla, Morelos, octubre de 1986**

| | *Positivas* | | *Negativas* | | |
|---|---|---|---|---|---|
| *Tipo* | *No* | *%* | *No* | *%* | *Total* |
| Hepatitis A<br>Anti hay Og G) | 60 | | 2 | 3.1 | 62 |
| Anti hay (Ig M) | 0 | 97.0 | 62 | 100.0 | 62 |
| Hepatitis B | | | | | |
| Anti Hbo Og G) | 5 | 9.4 | 57 | 90.6 | 62 |
| Anti Hbo Og M) | 0 | | 9 | 100.0 | 9 |
| HbsAg | 0 | | 61 | 100.0 | 61 |
| Anti Hbs | 2 | 33.3 | 4 | 66.6 | 6 |

* Insuficiente suero sanguíneo en un caso
Fuente: Laboratorio de Virologia, CDC, Atlanta.GA, EUA, 1986

## CONCLUSIONES

El estudio de brotes epidemiológicos es una disciplina de la epidemiología que recientemente ha recibido un importante impulso por parte de la comunidad epidemiológica. Por lo general, el diseño del estudio parte de una hipótesis inicial —controlando una gran variedad de factores— y de un evento epidemiológico agudo y de importante impacto en la salud pública, por lo que la celeridad en el accionar del epidemiólogo es de importancia esencial. Las herramientas con las que cuenta el estudio de brotes son, fundamentalmente, un investigador con una buena preparación epidemiológica que le asegure un actuar científico, con una buena y amplia formación médica, así como con la habilidad de socializar y capitalizar rápidamente las experiencias de los brotes, todo ello con el objetivo de identificar las formas de prevenir la transmisión futura de la enfermedad de que se trate.

El estudio presentado fue llevado hasta la identificación por microscopía electrónica del agente causal y permitió clasificarlo finalmente como virus de hepatitis E, de transmisión entérica. También, fue posible especificar las condiciones de saneamiento básico como los principales factores de riesgo para la transmisión y la persistencia de la enfermedad. Esto implica también aspectos culturales, hábitos, costumbres y condiciones materiales de vida de la comunidad estudiada.

Para resolver definitivamente este problema y evitar renuencias en comunidades con características similares, es importante impulsar la inversión para el desarrollo de pozos seguros o de plantas de tratamiento de agua que den servicio a los habitantes. Desafortunadamente este tipo de soluciones requiere de la inversión de recursos que generalmente no son accesibles para la comunidad y no representan una solución viable a corto plazo.

Se requiere de un kilo de madera para hervir un litro de agua durante un minuto; además, el agua se puede recontaminar aun después de hervida si se almacena en recipientes que no se han limpiado de manera apropiada. Otras opciones más costo-efectivas para desinfectar el agua incluyen la utilización de hipoclorito de sodio o de calcio, compuestos efectivos contra bacterias

**Cuadro VIII Algunas recomendaciones para reservorios de agua domiciliaria**

1. Usar recipientes translúcidos de plástico de polietileno de alta densidad, durables, ligeros, que no se oxiden, fáciles de limpiar y de bajo costo.
2. Que contengan volúmenes estándar (e.g., 20-L) con agarradera para su fácil manejo y transporte.
3. Un solo orificio de salida de 5 a 8 cm diámetro, con una cubierta fuertemente ajustada que permita el llenado fácil, así como el fácil ingreso de compuestos para la desinfección. También debe impedir la entrada de manos o utensilios para extraer agua.
4. Que incluya una boquilla durable y fácil de limpiar para extraer el agua.
5. La boquilla debe facilitar la entrada de aire junto con la salida de agua.
6. De preferencia, debe incluir marcadores de volumen e ilustraciones que garanticen un uso apropiado y seguro.

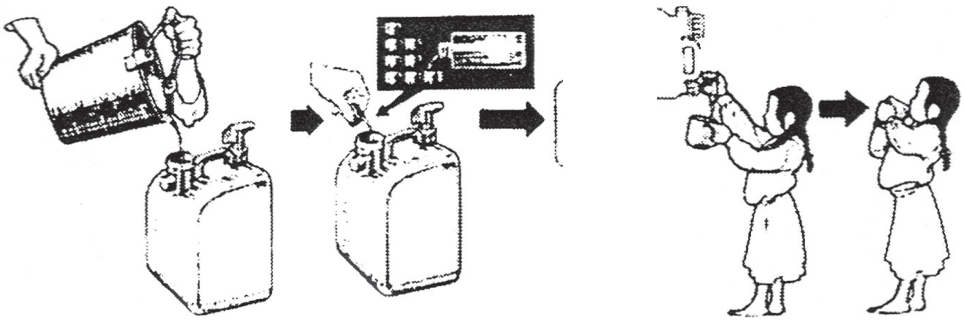
Fuente: Mintz ED, Reiff FM, Tauxe RV. Safe water treatment and storage in the home. A practical new strategy to prevent waterborn disease. JAMA 1995:273(12):948-953.

y virus; además, cuando se utilizan en recipientes cerrados, la protección puede durar por varios días (cuadro VIII).

En situaciones de emergencia se pueden usar blanqueadores y cloro comercial como el que se utiliza en el lavado de la ropa. Estos productos deben ser empleados con gran cautela, ya que es posible que contengan otros compuestos que pueden comprometer la salud.

Varios estudios realizados en diferentes partes del mundo han documentado que los contenedores utilizados para almacenamiento pueden también ser una fuente de contaminación para el agua. Diferentes investigaciones han documentado que los utensilios utilizados para servir agua o el contacto directo con las manos son fuente importante de contaminación. Se ha demostrado que el riesgo de contaminación es más alto cuando el agua es sacada con instrumentos o con las manos directamente, en contraste con la que se obtiene de recipientes con boca estrecha que obligan a servir directamente el agua (figura 2).





Fuente: Mintz ED, Reiff FM, Tauxe RV. Safe water treatment and storage in the home. A practical new strategy to prevent waterborn disease. JAMA 1995:273(12):948-953.

**Figura 2 Ilustración para llenar, desinfectar y utilizar apropiadamente un reservorio de agua para consumo intradomiciliario**

## RESPUESTAS

1. Una epidemia se refiere a un aumento en la frecuencia con la que se presenta una enfermedad o evento de salud. Operacionalmente, las epidemias se identifican comparando la frecuencia actual de un evento o enfermedad —ya sea esta última crónica o infecciosa— con la que habitualmente se presenta. Así, cuando se manifiesta un incremento en la frecuencia habitual de casos de una enfermedad, puede decirse que existe una epidemia.
2. Con la información disponible no se puede determinar si este evento constituye una epidemia, y para poder hacerlo es necesario comparar la frecuencia actual con la que se ha registrado en años anteriores. Es preciso saber la tasa de incidencia habitual, es decir; el número de casos de ictericia que uno esperaría encontrar en ese lugar; en ese tiempo y en esa estación del año. Esto se puede estimar por el número de casos de ictericia notificados durante el año anterior. Si el número de casos encontrados es superior comparado con lo esperado, se debe descartar que el aumento sea consecuencia de un artefacto resultante de cambios que pudieron haber ocurrido en el sistema de vigilancia (por ejemplo, un médico nuevo, un oficial nuevo de salud pública, o bien, innovaciones en pruebas diagnósticas, laboratorios o lineamientos de comunicación de caso). Si no se cuenta con registros epidemiológicos —lo que es común en localidades rurales pequeñas—, se puede utilizar información derivada de entrevistas hechas con los médicos, el personal de salud o los curanderos que dan servicio a la comunidad; ello contribuirá a establecer la frecuencia habitual de este padecimiento.
3. Es poco usual un informe de 31 casos de enfermedad ictérica (hepatitis) en una comunidad rural de menos de 2 000 habitantes, particularmente con las características clínicas descritas; esto se debe a que es raro que se notifiquen brotes de hepatitis en adultos de países como México, donde la hepatitis A es endémica y afecta desde muy temprana edad a la población.
4. En el cuadro IX se presenta una guía general para realizar un estudio en el que se sospecha que la etiología está relacionada con la contaminación del agua.
5. De acuerdo con el informe inicial hecho por el médico de la comunidad, el elevado número de personas que fueron afectadas sugiere una fuente común de transmisión como el agua o los alimentos. Adicionalmente, la observación que indica que la mayoría de los casos ocurrieron en la población adulta, también sugiere la consideración de posibles fuentes ocupacionales o sociales como podría haber sido alguna reunión a la que haya acudido un porcentaje alto de los adultos de la comunidad. Esto último también es una posibilidad, ya que ambos sexos fueron afectados con igual frecuencia. En general, las intoxicaciones de origen ocupacional tienen una mayor repercusión en el grupo de hombres. Las posibilidades diagnósticas que se deben descartar incluyen:

   Hepatitis A. Ocurre comúnmente en grupos de población de todo el mundo; su principal ruta de transmisión es oro-fecal y de persona a persona, o por agua o alimentos contaminados. Se han notificado numerosos casos de epidemias causadas por el virus de la hepatitis A después de la ingesta de agua o alimentos contaminados con este virus. En países como México, la hepatitis A es endémica, y la mayoría de la población entra en contacto con el virus a edades tempranas, lo que resulta en una infección subclínica. Diferentes encuestas indican que más de 90% de la población de 40 años tiene anticuerpos contra este

**Cuadro IX Guía de acciones para investigar una posible epidemia por contaminación de agua**

1. Determine la existencia de la epidemia y verifique el diagnóstico:
   a) Obtenga historias clínicas de los casos.
   b) Obtenga muestras biológicas relevantes de los casos y rnuestras de agua.
   c) Desarrolle una definición operacional de caso.
2. Cuente los casos.
3. Describa la ocurrencia de los casos en términos de tiempo, espacio y persona.
4. Determine quién está en riesgo de padecer el evento o enfermedad.
5. Proponga hipótesis para determinar el o los mecanismos de transmisión/infección que hayan contribuido al desarrollo de la epidemia.
6. Compare sus hipótesis con los datos recolectados en el campo y con otras fuentes de información.
7. Desarrolle un estudio analítico a fin de probar las hipótesis propuestas.
8. Elabore un informe escrito donde se documenten todas las acciones.
9. Prepare medidas para controlar la epidemia.
10. Prepare y describa medidas de control para evitar brotes futuros.
11. Inicie programas para garantizar el aporte de agua segura.

virus. Los casos secundarios se desarrollan entre 2 y 6 semanas después del contacto con el caso infectado.

Hepatitis E (anteriormente llamada hepatitis no A no B). Se le ha notificado en brotes de grandes proporciones y la mayoría de los casos con sintomatología clásica de hepatitis (ictericia, coluria, malestar general) ocurren en adultos jóvenes. La baja tasa de ataque que se presenta en niños probablemente se deba a que en este grupo se presentan formas subclínicas. El virus de la hepatitis E es similar al de la hepatitis A y se transmite por contaminación fecal del agua.

Hepatitis B, E y D. Se presenta en habitantes de todo el mundo, pero el principal mecanismo de transmisión es el contacto con la sangre de personas infectadas, ya sea por transfusiones, inyección de drogas intravenosas o punciones accidentales en hospitales. El virus de la hepatitis B también se puede transmitir por contacto sexual y por vía perinatal. Es raro que este tipo de hepatitis se presente con carácter epidémico en comunidades rurales.

Leptospirosis. La presencia de este padecimiento es imposible, porque faltan signos relacionados con las meninges, daños renales o erupción de la piel. El periodo de incubación es corto (de 4 a 19 días). Las fuentes de transmisión pueden ser animales salvajes o domésticos, y en casos muy raros, otros humanos. El mecanismo de transmisión es el contacto con la orina de animales infectados, con caminos o ríos contaminados, o bien, con los cuerpos de animales muertos.

Intoxicaciones por fósforo, arsénico, plaguicidas, hongos o solventes orgánicos. Es conveniente considerar la exposición a plaguicidas; sin embargo, es improbable una intoxicación masiva que se acompañe de fiebre, además de que la severidad sería mayor, incluyendo la letalidad. En este caso no había historia de exposición a toxinas de origen vegetal. No obstante, puede ser difícil diferenciar entre hepatitis viral y hepatitis tóxica, solamente con datos clínicos o del laboratorio.

Reacciones a fármacos. Las sustancias de uso médico que pueden provocar cuadros hepáticos con ictericia se agrupan como: antifímicos, antirreumáticos, anticonvulsivantes, y hormonas terapéuticas. Sin embargo, estos casos se presentan de manera aislada y eventual. Es muy improbable la presentación epidémica de hepatitis por estas causas, y no existe en ningún caso el antecedente de exposición a este tipo de sustancias. Es importante tener en cuenta el diagnóstico probable, ya que el análisis gráfico de la ocurrencia del evento en el tiempo, recomienda la creación de estratos o ventanas de tiempo con una duración de 1/3 a 1/4 del periodo de incubación de la posible enfermedad.

6. Para este estudio en particular, se definió como caso cualquier persona que hubiese tenido un inicio abrupto de ictericia, con residencia en la comunidad de estudio, y que hubiese comenzado a presentar signos y síntomas entre el 1 de junio y el 29 de octubre de 1986.

   Es necesario enfatizar que una definición de caso debe partir de criterios más abiertos y ser dinámica, y que puede requerir refinamientos cuando se obtenga más información. Generalmente incluye criterios de tiempo, lugar y persona.
7. No, es muy probable que con esta definición operacional no se haya identificado la totalidad de casos de enfermedad. Es posible que no se haya identificado a todos los sujetos que cursaron con formas leves de la enfermedad, las cuales se acompañan de ictericia tenue o subclínica. Ciertamente, la definición operacional excluye a los sujetos que cursaron con un cuadro de hepatitis anictérica. La definición operacional se podría haber completado con la combinación de algunos síntomas clínicos: sin embargo, se consideró utilizar únicamente la presencia de ictericia, ya que ésta es un indicador específico del cuadro y puede ser identificada fácilmente por la comunidad y los médicos o personal de salud que dan servicio a la comunidad. Adicionalmente, se podría haber recurrido a otras técnicas diagnósticas como las pruebas de enzimas hepáticas o de bilirrubina en la sangre de las personas con sospecha de enfermedad: no obstante, dadas las condiciones de la comunidad y la falta de laboratorios o infraestructura médica accesible, se decidió utilizar una definición de caso simple y fácil de operar con los recursos disponibles en el momento de la investigación.
8. Se sugiere lo siguiente:
   a) Entrevistar a los casos identificados y a sus familiares, y preguntarles si conocen a otras personas que hayan padecido un cuadro de ictericia durante el periodo de estudio.
   b) Hacer contacto con médicos y otros profesionales de la salud, así como con hospitales y clínicas que dan servicio a la comunidad.
   c) Revisar los registros del departamento de salud local.
   d) Establecer una vigilancia activa (búsqueda intencionada de casos), mediante visitas domiciliarias o realizando un censo de la comunidad.
   e) En otras situaciones, se podría considerar el uso de la prensa, la radio y la televisión para estimular la notificación de casos.
9. En las figuras 3 y 4 se muestran las curvas epidémicas de la enfermedad ictérica entre los residentes de Huitzililla, Morelos, de acuerdo con la fecha de inicio de los síntomas y con la semana de diagnóstico. La epidemia alcanzó un pico a finales de agosto. En este mes se diagnosticó 44% de los casos, y el último caso correspondiente al brote inicial se diagnosticó el 29 de octubre.

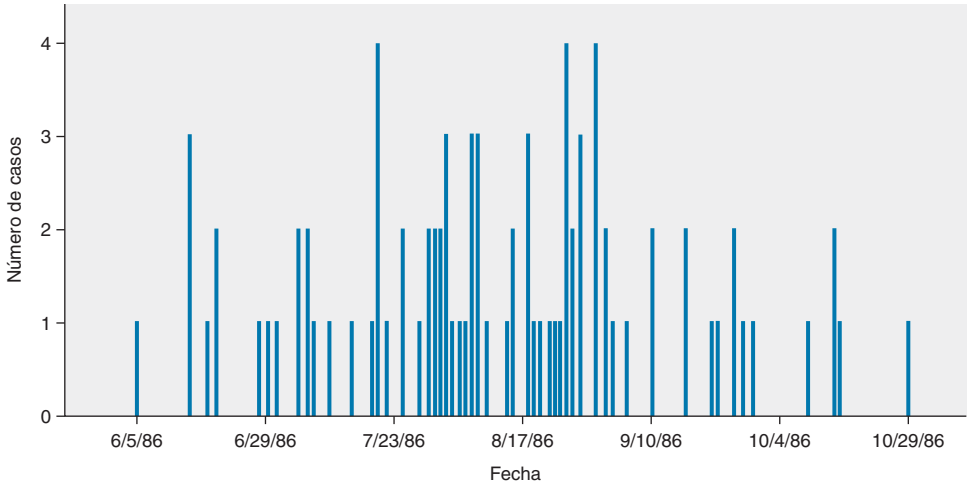



**Figura 3** **Distribución de los casos de enfermedad ictérica de acuerdo con la fecha de diagnóstico. Residentes de Huitzililla, Morelos, junio-octubre de 1986**



Para la interpretación de la curva epidémica es necesario tener en cuenta los siguientes aspectos: los datos deben representarse como un histograma en el que el tiempo de interés abarque la duración de la epidemia. El tiempo debe dividirse en intervalos de duración que correspondan a 1/3 o 1/4 del tiempo de incubación; esto es particularmente importante cuando se sospecha que existe transmisión de persona a persona. Se debe examinar la forma de la curva epidémica y caracterizar los puntos de máxima incidencia o tasa de ataque. Siempre que sea posible o relevante, se debe construir la curva epidémica de acuerdo con variables de interés, por ejemplo grupos de edad, género, lugar de residencia, etcétera.

El estudio de la curva epidémica es importante para estimar la magnitud del problema y generar hipótesis. Una epidemia que muestra un solo pico máximo de incidencia sugiere una fuente común única, por ejemplo: una bebida o un alimento contaminados. Una curva epidémica con un punto máximo y una caída lenta o prolongada en el tiempo sugiere un agente que se transmite de persona a persona. Cuando la epidemia muestra varios picos o una tasa de ataque sostenida se debe sospechar que existe una fuente continua o permanente. Por ejemplo, las epidemias originadas por insectos pueden generar este tipo de curva epidémica. En el caso que nos ocupa, la curva epidémica sugiere una fuente común que posteriormente se acompañó de una transmisión de persona a persona, lo cual concuerda con la hipótesis de que se trata de una posible epidemia de hepatitis y sugiere que los casos que no son de fuente común se explican por dicha transmisión.

Se puede estimar que la exposición ocurrió en algún momento durante la primera semana de junio si se considera que, en los primeros días de ese mismo mes, coincidían los

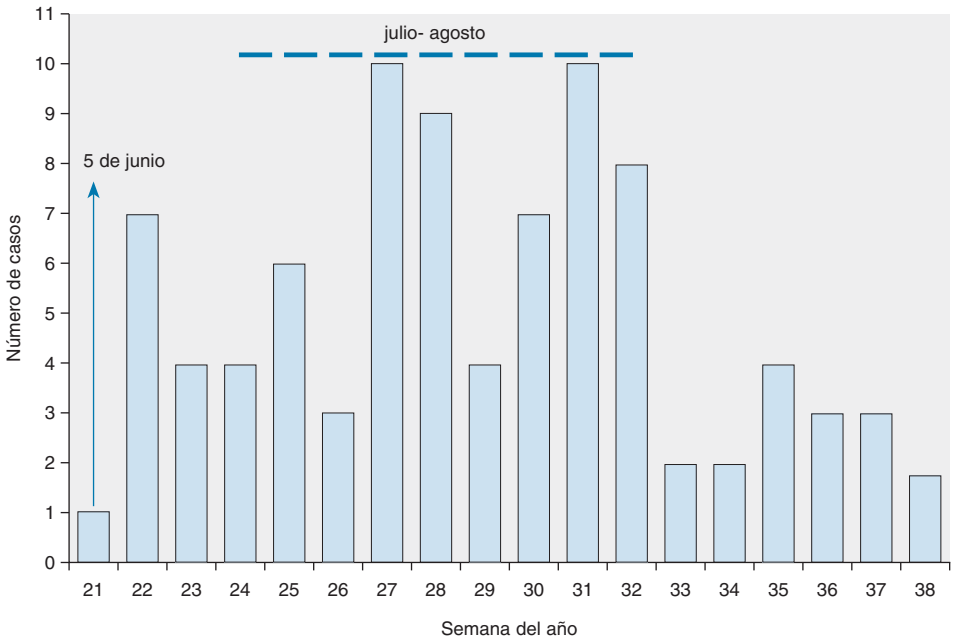





**Figura 4** **Distribución de los casos de enfermedad ictérica de acuerdo con la semana de diagnóstico. Residentes de Huitzililla, Morelos, junio-octubre de 1986**

cuarenta días anteriores al pico del brote (incubación promedio para hepatitis E). Para poder decidir sobre la existencia de una epidemia es necesario considerar la frecuencia habitual de la enfermedad en esa comunidad.

10. Cuando la epidemia es identificada en sus etapas iniciales, las pruebas para buscar patógenos en el agua pueden ser de utilidad para determinar el agente causal. Sin embargo, es necesario considerar que el agente puede presentarse en concentraciones muy bajas, lo que dificultaría su documentación con la tecnología disponible para el estudio de estos casos en campo. Se puede medir la presencia de coliformes como un indicador de contaminación fecal. Algunas consideraciones sobre los límites de coliformes se presentan en el cuadro X.
11. No, se requiere de realizar pruebas diagnósticas. Se podrían utilizar pruebas serológicas para documentar la presencia de anticuerpos del virus de la hepatitis o, alternativamente, buscar partículas virales en muestras de materia fecal de casos en fase aguda.
12. La tasa de ataque se puede estimar dividiendo el número de casos de enfermedad ictérica entre la población en riesgo. La tasa de ataque varía de acuerdo con la localización geográfica; los casos se distribuyeron en 17 de las 20 manzanas que conforman la localidad. Las tasas de ataque por manzana variaron desde 0 hasta 22.6%. Al analizar la infor-

**Cuadro X Guía para la interpretación de la calidad de agua en situaciones de desastre**

| *Coliformes por 100 ml de agua* | *Calidad del agua* |
|---|---|
| 0-10 | Razonable |
| 10-100 | Contaminada |
| 100-1000 | Peligrosa |
| > 1000 | Muy peligrosa |

mación se encontró que en cuatro manzanas se alcanzaron tasas de 7.7 a 23%; en nueve, las tasas observadas fueron de 3.4 a 7.7%; en cinco manzanas, de 0.62 a 3.3%, y hubo dos manzanas sin casos (cuadro XI).

Según la edad, las tasas de ataque más altas se presentaron en quienes tienen de 15 a 24 años (9.6 × 100), y de 25 a 44 años (8.6 × 100) (cuadro XII). De acuerdo con el sexo, las tasas de ataque en población masculina fueron: 4.9 × 100 = 44 casos, y en población femenina, 5.1 × 100 = 44 casos, total = 88 casos (cuadro XIII). Por lo tanto, 86.4% de los casos son mayores de 15 años, y la relación de casos por sexo es de 1:1. Se puede ob-

**Cuadro XI Casos de enfermedad ictérica y tasas de ataque por 100 habitantes, según manzana de residencia. Huitzililla, Morelos, 1986**

| *Manzana* | *Número de habitantes* | *Número de habitantes* | *Tasa de ataque por 100 habitantes* |
|---|---|---|---|
| 14 | 0 | 70 | 0 |
| 20 | 0 | 47 | 0 |
| 16 | 1 | 160 | 0.63 |
| 13 | 1 | 81 | 1.23 |
| 3 | 1 | 76 | 1.32 |
| 17 | 1 | 75 | 1.33 |
| 2 | 1 | 46 | 2.17 |
| 11 | 6 | 173 | 3.47 |
| 18 | 3 | 86 | 3.49 |
| 19 | 4 | 111 | 3.6 |
| 10 | 3 | 75 | 4.0 |
| 4 | 5 | 110 | 4.55 |
| *15 | 11 | 228 | 4.82 |
| *6 | 5 | 102 | 4.9 |
| 7 | 1 | 18 | 5.56 |
| 8 | 3 | 50 | 6.0 |
| 9 | 4 | 52 | 7.7 |
| *5 | 19 | 113 | 16.81 |
| *1 | 12 | 62 | 19.35 |
| *12 | 7 | 31 | 22.6 |
| Totales | 88 | 1757 | 5.01 |

* Manzanas cercanas a los arroyos

**Cuadro XII** **Casos de enfermedad ictérica, población en riesgo y tasas de ataque por 100 habitantes de acuerdo con la edad. Huitzililla, Morelos, 1986**

| *Grupo de edad* | *Número de casos* | *Población en riesgo* | *Tasa de ataque por 100 habitantes* | *Razón de tasa (IC95%)* |
|---|---|---|---|---|
| Menos del año | 0 | 62 | 0.0 | 0 |
| 1-4 | 3 | 235 | 1.28 | 1.0 |
| 5-14 | 9 | 583 | 1.54 | 1.2 (0.33-4.43) |
| 15-24 | 32 | 333 | 9.61 | 7.53 (2.33-24.29) |
| 25-44 | 30 | 348 | 8.62 | 6.75 (2.08-21.87) |
| 45 y más | 14 | 196 | 7.14 | 5.59 (1.63-19.19) |

servar que la tasa de ataque —tomando como referencia el grupo de menores de cuatro años— aumenta conforme se incrementa la edad, para disminuir lentamente a partir de los 24 años.

Se observó una marcada concentración de casos alrededor de dos pequeños arroyos, Chapala y Venero del Sabino, que bordean la localidad. Ahí se encontraron tasas de ataque de 19.35, 16.81 y 4.9 para las manzanas 1, 5 y 6, respectivamente, y de 22.6 y 4.82 para las manzanas 12 y 15, en el mismo orden. Estas manzanas se encuentran a la orilla de estos cursos de agua (cuadro XI). De igual forma, el riesgo fue más evidente en las cercanas a los arroyos (cuadro XIV).



13. Como primer paso, es necesario hacer un resumen de la epidemiología descriptiva como el siguiente:
    13.1. La curva epidémica sugiere un brote de hepatitis con fuente común y continua (por transmisión de persona a persona y muy probablemente por exposición continua al agua contaminada) y, además, que la exposición inicial ocurrió durante algún tiempo en las primeras semanas de junio.
    13.2. El brote se originó en los arroyos de la localidad.
    13.3. El brote afecta en mayor grado a adultos jóvenes de 15 a 24 años.

A partir de ello, se pueden hacer las siguientes hipótesis respecto a la fuente de infección y el modo de transmisión:

**Cuadro XIII** **Casos de enfermedad ictérica, población en riesgo y tasas de ataque por 100 habitantes de acuerdo con el sexo. Huitzililla, Morelos. 1986**

| Sexo | *Número de casos* | *Población en riesgo* | *Tasa de ataque por 100 habitantes* |
|---|---|---|---|
| Masculino | 44 | 900 | 4.9 |
| Femenino | 44 | 857 | 5.1 |

**Cuadro XIV** **Casos de enfermedad ictérica, población en riesgo y tasas de ataque de acuerdo con la localización de la residencia en relación con los arroyos. Huitzililla, Morelos, 1986**

| | *Cerca* | *Lejos* | *Total* |
|---|---|---|---|
| Casos | 54 | 34 | 88 |
| Población en riesgo | 536. | 1221 | 1757 |
| Tasa de ataque | 0.100 | 0.027 | 0.047 |
| | Estimador | [IC95%] | |
| Diferencia de tasas | 0.07 | 0.04 a 0.10 | |
| Razón de tasas | 3.61 | 2.31 a 5.73 | |

a) Considere los mecanismos posibles de transmisión: persona a persona, probable fuente común, probable comida (la que se preparó en el restaurante local, la tienda, la iglesia, ostras, almejas, leche, agua).
b) Hasta ahora sabemos que es imposible que la transmisión sea exclusiva de persona a persona. Por lo tanto, nuestra hipótesis se enfoca en una transmisión de origen hídrico. Se descartan las de origen alimentario como fuente común, porque cerca de los dos arroyos se concentran las tasas de ataque más altas. También es preciso considerar que adultos jóvenes de 15 a 24 años tenían las tasas de ataque más altas. Cuando se investigan brotes, es importante buscar en dónde hay más casos, es decir, caracterizar los grupos de alto riesgo para generar una hipótesis.

14. Para descartar o apoyar las diferentes hipótesis, se necesita información adicional sobre el uso del agua, las prácticas higiénicas y el estado de los pozos de agua domiciliarios que se utilizan para consumo humano. Igualmente se requiere de información adicional para descartar otras hipótesis como la referente a la contaminación de alimentos. En esta etapa de la investigación, el equipo de trabajo se dio a la tarea de desarrollar un cuestionario que permitiera evaluar y contrastar las hipótesis planteadas, así como incluir marcadores biológicos que permitieran establecer con mayor certeza el agente etiológico.
15. Con el fin de obtener información más específica de los posibles factores de riesgo, se podría estudiar una muestra de los casos que ocurrieron en la primera fase del brote y comparar las características de éstos con los de una muestra de personas de la misma edad y sexo de la población que no presentó el evento en estudio (controles).

    En este caso es importante que los controles se seleccionen de entre aquellas personas que estuvieron en riesgo de presentar el evento y no lo padecieron. En este estudio, dado que los recursos y los tiempos eran limitados, se decidió estudiar únicamente 32 casos y 20 controles. En general, lo recomendable es estudiar el número necesario de casos y controles que permitan detectar con cierta precisión y confiabilidad las asociaciones que se presentan como hipótesis de trabajo.
16. Véase el cuadro XV.

**Cuadro XV Guía para la inspección de pozos de agua para consumo humano**

1. ¿El área de extracción del agua se encuentra protegida?
2. ¿Cómo es el sistema de extracción de agua?
3. ¿El sitio donde está el pozo puede ser alcanzado durante inundaciones?
4. ¿El sitio del pozo está cerca de alguna fuente de contaminación?
5. ¿Tiene recubierta en la pared interna?
6. ¿El pozo está bien sellado?
7. ¿Es accesible a la población?

17. Al analizar la información de casos y controles para la identificación de factores de riesgo asociados con la presencia de hepatitis, se observó que la utilización de agua hervida tiene efectos protectores contra la enfermedad ($p<0.01$). Del mismo modo, la protección interna de los pozos familiares con revestimiento de concreto demostró dicho efecto protector ($p<0.01$). En contraste, la presencia de agua sucia o contaminada en el pozo familiar estuvo fuertemente asociada con la presencia de la enfermedad ($p=0.01$); asimismo, el contacto con casos de hepatitis representó un riesgo significativo de enfermar ($p<0.01$) (cuadro VI). En otras variables, identificadas inicialmente como posibles factores de riesgo, no se encontraron asociaciones estadísticas significativas (por ejemplo, alimentos, relaciones sexuales, aplicación de inyecciones, etc.).



Para estimar la razón de momios, el intervalo de confianza y el valor *p* se pueden utilizar las siguientes fórmulas;

| | Hierve el agua | | |
|---|---|---|---|
| | Sí | No | |
| Casos | a = 7 | c = 25 | N1 = 32 |
| Controles | b = 10 | d = 10 | N0 = 20 |
| | $M_1 = 17$ | $M_0 = 35$ | T = 52 |

$$RM = \frac{a \times d}{b \times c} = \frac{7 \times 10}{25 \times 20} = 0.28$$

$$\chi^2 = \frac{T\left[(a \times d) - (b \times c)\right]^2}{M_1 \times M_0 \times N_1 \times N_0} = \frac{1684800}{380800} = 4.42$$

$$\chi^2 \,|g|\, 4.42 = 0.0.35$$

En este caso se estimó el valor de $X^2$ y se comparó con la tabla de probabilidades para obtener un valor de 0.035. Asimismo, el valor *p* puede interpretarse como la probabili-

dad de obtener un resultado tan extremo o mayor al observado bajo la suposición de que la hipótesis nula es cierta.

Para estimar el intervalo de confianza se utilizó la aproximación normal, que implica, la transformación logarítmica con la siguiente fórmula:

$$\ln(RM) \pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \ln(0.28) \pm 1.96\sqrt{\frac{1}{7}+\frac{1}{25}+\frac{1}{10}+\frac{1}{10}} =$$

$$= \exp(-1.27 \pm 1.21) = (0.083,\ 0.94)$$

18. Se requiere demostrar la presencia de marcadores biológicos de exposición al agente, o bien, del agente etiológico en heces de los casos que se encuentran en etapa aguda.
19. Los datos sugieren que no se trata de hepatitis A, dado que todos los sujetos tenían anticuerpos contra este tipo de virus (lgG) y no se detectaron anticuerpos de tipo IgM que indicarían una exposición reciente al virus. Los datos también descartan que se trate de hepatitis tipo B. Es necesario recuperar información bibliográfica internacional que permita comparar las características de este brote con otros, buscando semejanzas o correlaciones.

## BIBLIOGRAFÍA


Alter MJ *et al.* Sporadic non-A, non-B hepatitis. frecuency and epidemiology in an urban U.S. population. lnfect Dis 1982;145:6.

Belabbes EH. Bourguermouh A, Benatallah A, Illoul G. Epidemic non A, non B viral in Algeria: strong evidence for its spreading by water. J Med Virol 1985; 16:257-263.

Bay O. Hadler S. Maynard J. La hepatitis en las Américas. Boletín Epidemiológico OPS 1985:6-5.

Benenson AS. Epidemic non-A, non-B hepatitis. En: Control of comunicable diseases in man.Washington, D.C.: The American Public Health Association. 1985: 179-181.

Hepatitis virica en la región. Bol Oficina Sanit Panam 1989; 100:3.

Bradley DW, Andjaparidze A, Cook EH Jr, McCaustland K, Balayan M, Stetler *et al.* Aetiologica1 agent oftransmitted non-A, non-B hepatitis. J Gen ViroI 1988;69:731-738.

Bradley DW, Maynard JE. Etiology and natural history ofpost-transfusion andentere cally transmitted non-A non-B hepatitis. Semin Liver Dis 1986;6:1.

Bustamante CM *et al.* Etiología de la hepatitis en la ciudad de México. Bol Med Hosp Infant Mex 1986:43;41.

Calderón JE *et al.* Hepatitis infecciosa. Presencia del antígeno HBs. Bol Med Hosp Infant Mex 1975:42:6.

Estadísticas de la Dirección General de Epidemiología. Dirección de Información y Cómputo, 1986.

Joshi YK, Babu S, Sarin S, Tandon BN, Gandhi BM, Chaturvedi VC. Inmunoprofilaxis de la hepatitis no A no B epidémica. Indian J Med Res 1985;81:18-19.

Kane MA, Bradley DW, Shrestha SM, Maynard JE, Cook EH, Mishra RP *et al.* Transmission studies in marmosets. JAMA 1984:252(22):3140-3145.

Khuroo MS. Estudio de una epidemia de hepatitis no A, no B. Am J Med 1980:68.

Khuroo MS, Duermeyer W, Zargar SA, Ahanger MA, Shah MA. Acute sporadic non-A, non-B hepatitis in India. Am J Epidemiol 1983; I 18:360-364.

Kumate J, Alviwuri AM, Isibasi A. Encuesta serológica de hepatitis A en niños de México. Oficina Sanit Panam 1982;92:494.

Masayoshi Y, Nakajima H, Kimura K, Fujisawa K, Kameda H, Nakahara M *et al.* An epidemic ofnon-A, non-B hepatitis in Japan. Am J Gastroenterol 1983;78(10):652-655.

Maynard JD. Hepatitis no A no B epidémica. Semin Liver Dis 1984;4:4.

Myint Myint Soe, Tun Khin, Thein-Maung Myint, Khin Maung Tin. A clinical and epidemiological study of an epidemic of non-A, non-B hepatitis in Rangoon. Am J Trop Med Hyg 1985;34: 1183-1189.

Nagata A, Kiyosawa K, Koike Y, Miura M, Gibo Y, Sodeyama T *et al.* Epidemiology of sporadic acute non-A, non-B hepatitis in Japan: A comparison with hepatitis A and B. Am J Gastroenterol 1985;80(4);298-302.

Ruiz GJ. Anticuerpos de hepatitis A. Prevalencia en un grupo de niños mexicanos. Am J Epidemiol 1985;21:1.

Sreenivasan MAJ. Estudio seroepidemiológico de hepatitis viral en Kolhapur, India. Indian J Med Res 1984; 93; 113.

Tandon BN, Joshi YK, Jain SK. Epidemia de hepatitis no A no B en el norte de India. Indian J Med Res 1982;75:739.

Velázquez O, Stetier HC, Ávila C, Ornelas G, Alvarez C, Hadler SC *et al.* Epidemic transmission of enterically transmited Non-A, Non -B hepatitis in México. 1986-1987. JAMA 1990;263:3281-3285.

Wong DC, Purcell RH, Sreenivasan MA, Prasad SR, Pavri KM. Epidemic and endemic hepatitis in India, evidence for a non-A, non-B hepatitis virus aetiology. Lancet 1980;2:876-879.

# X Sesgos

*Mauricio Hernández Ávila*
*Francisco Garrido Latorre*
*Eduardo Salazar Martínez*

Un requisito fundamental de cualquier estudio epidemiológico que desee conocer la frecuencia con la que ocurre un evento —o cuantificar la asociación entre un factor de riesgo y una enfermedad— es estimar con la mayor validez y precisión posible sus determinaciones. En otras palabras, la exactitud del conocimiento derivado de un estudio epidemiológico dependerá, en gran medida, de la ausencia de error y la capacidad de estimar o predecir el parámetro verdadero en la población blanco. En el contexto de la epidemiología, un estudio puede considerarse válido si no presenta sesgos o errores sistemáticos. A lo largo del presente capítulo se hará referencia a dos tipos de validez: a) la validez interna, que se refiere principalmente a la ausencia de sesgos cometidos durante la selección de la población de estudio, las mediciones realizadas en ella o el momento de asegurar la comparabilidad de los grupos estudiados, y b) la validez externa, que se refiere a la capacidad del diseño efectuado para generalizar sus resultados observados en la población fuente de estudio a otras poblaciones, lugares y momentos diferentes de los estudiados, hasta convertirse en una asociación plausible. Es importante hacer notar que la validez interna es una condición necesaria para evaluar la presencia de validez externa; es decir, es necesario primero cumplir con los requisitos de validez interna para entonces poder generalizar los resultados, que es el propósito de la validez externa. Por esta razón, los estudios epidemiológicos privilegian acciones que maximizan la validez interna, aun comprometiendo, en cierta medida, la validez externa.

Todo estudio epidemiológico, particularmente los observacionales, está sujeto a un cierto margen de error, por lo que es importante conocer sus fuentes principales y los diferentes procedimientos que pueden utilizarse para minimizar su impacto en los resultados. El error, desde el punto de vista epidemiológico, se clasifica en dos tipos: los sesgos (también conocidos como errores sistemáticos), y los errores aleatorios. Ambos tipos de error, si no se controlan adecuadamente, pueden comprometer la precisión y validez del estudio. La falta de precisión ocurre cuando las mediciones repetidas, ya sean en un mismo sujeto o en diferentes miembros de la población en estudio, varían de manera impredecible, mientras que el sesgo ocurre cuando estas medidas varían de manera predecible y, por lo tanto, provocan una tendencia a sobre o subestimar el valor verdadero en medidas repetidas.

La analogía utilizada para describir ambos conceptos es la práctica de "tiro al blanco", donde la diana es el valor verdadero en la población blanco, y los "disparos" son las diferentes mediciones que se realizan para estimar dicho valor verdadero. Un buen tirador, cuya arma no esté bien calibrada, podrá ser muy preciso y lograr que todos los disparos den en el

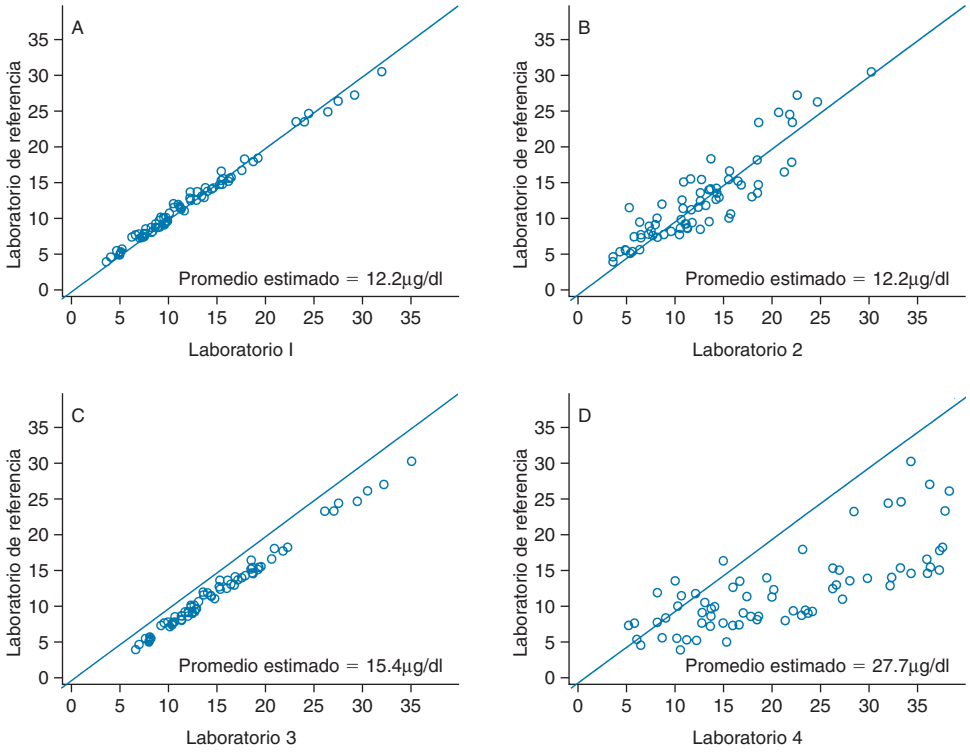



**Figura 1** **Valores de plomo en sangre reportados en cuatro laboratorios con diferentes condiciones de precisión y validez. La línea diagonal refleja el valor real de la medición**



mismo lugar; pero ninguno de ellos dará en el blanco correcto. Esto corresponde al sesgo. Por otra parte, un tirador con un arma bien calibrada, pero con mano temblorosa, podrá apuntar hacia el blanco correcto, pero sus disparos no darán en la diana.

Puede aplicarse la analogía descrita en el contexto de un estudio epidemiológico, cuyo principal objetivo sea determinar las concentraciones de plomo en la sangre de la población general. Supóngase que en este estudio participan cuatro laboratorios y un laboratorio de referencia, en el cual se repite cierto porcentaje de las determinaciones de plomo que llevan a cabo los otros laboratorios. Los resultados hipotéticos se muestran en la figura 1. En el panel A se muestra un laboratorio cuyas determinaciones son válidas y precisas respecto de las mediciones que realiza el laboratorio de referencia (situación ideal), ya que el valor estimado corresponde muy cercanamente al valor real; en el panel B se muestran resultados de un laboratorio cuyas mediciones son válidas, pero con precisión intermedia; en el laboratorio del panel C la validez es cuestionable, pero la precisión es aceptable; mientras que en el panel D, las mediciones son imprecisas y no válidas.

Mientras que en los dos primeros escenarios el parámetro estimado en la población en estudio es el mismo —en ambos laboratorios se estima una media poblacional de 12.2 μg/dl, que corresponde a la media verdadera de 12.2 μg/dl—, en los páneles C y D se estimaría una media de 15.4 μg/dl y 27.7 μg/dl, respectivamente. La utilización de laboratorios con resultados como los que se muestran en los páneles C y D arrojaría resultados claramente erróneos al sobrestimar el valor real, lo que nos llevaría a concluir equivocadamente que existen concentraciones de plomo en sangre más elevadas que las reales.

En los estudios epidemiológicos analíticos, en los que se pone a prueba una hipótesis por medio de la comparación de dos o más grupos de estudio, los errores en la estimación de la asociación también pueden ser tanto aleatorios o no sistemáticos como sistemáticos. El error sistemático o sesgo se ha definido como cualquier error que ocurre de manera diferencial —en relación con los grupos que se comparan— y que tiene origen en el diseño, conducción o análisis del estudio, y que invariablemente resulta en una conclusión errónea, ya sea proporcionando una estimación más baja o más alta del valor real de la asociación que existe en la población blanco.

Dependiendo de la etapa del estudio en que se originan, los sesgos que interfieren con la validez interna de un estudio se han clasificado en tres grandes grupos: a) de selección, que se refieren a los errores que se introducen durante el proceso de selección o el seguimiento de la población en estudio; b) de información, que son errores en la estimación de la asociación en los que se incurre durante los procesos de medición en la población en estudio; y c) de confusión, que básicamente se originan ante la falta de comparabilidad de los grupos estudiados que resulta cuando no es posible asignar en ellos la exposición de manera aleatoria. Todo diseño de investigación epidemiológica, en mayor o menor medida, es susceptible de este tipo de sesgos, por lo que es un imperativo para los investigadores planear adecuadamente cada una de las etapas de un estudio con el propósito de evitar, o disminuir al máximo, la probabilidad de incurrir en dichos errores durante la estimación de la asociación. Sin duda, la etapa más crítica de un estudio corresponde al planteamiento del diseño, ya que resulta casi imposible corregir a posteriori los sesgos introducidos durante esta etapa.



## SESGOS DE SELECCIÓN

El sesgo de selección es considerado un error sistemático en la estimación de la asociación que se introduce durante el proceso de selección o el seguimiento de la población en estudio y que propicia una conclusión equivocada sobre la hipótesis que se evalúa. Los sesgos de selección pueden ser originados por el mismo investigador o ser resultado de relaciones complejas en la población del estudio que no son evidentes para el investigador y pasan desapercibidas. En este contexto, una posible fuente de sesgo de selección puede ser cualquier factor que influya sobre la probabilidad de los sujetos seleccionados de participar o permanecer en el estudio y que, además, esté relacionado con la exposición o con el evento en estudio.

Los sesgos de selección pueden ocurrir en cualquier estudio epidemiológico; sin embargo, ocurren con mayor frecuencia en estudios retrospectivos y, en particular, en estudios transversales o de encuesta. En los estudios de cohorte prospectivos los sesgos de selección

ocurren con menor frecuencia, ya que el reclutamiento y selección de la población en estudio se da antes de que ocurra el evento que se estudia, así que puede suponerse que la selección de los participantes se realiza independientemente del evento y, en general, la participación en el estudio no puede ser influida de manera directa por el evento, ya que éste aún no ha ocurrido. En contraste, la permanencia de los participantes en el estudio sí puede ser determinada por el evento. Cuando esto ocurre, y es de diferente magnitud para los grupos expuesto y no expuesto, existirá la posibilidad de que los resultados se vean distorsionados por esta permanencia diferencial. Por esta razón, se recomienda maximizar las tasas de permanencia y seguimiento en los estudios de cohorte.

En los estudios retrospectivos, los sesgos de selección pueden ocurrir cuando los participantes potenciales o los investigadores conocen la condición de exposición y/o de evento, y este conocimiento influye diferencialmente en la participación en el estudio. En el contexto de este tipo de estudios, cualquier factor que influya sobre la probabilidad de selección, ya sea como caso o como control, y que a su vez esté relacionado con la exposición en estudio, será una posible fuente de sesgo de selección. Supóngase que se llevará a cabo un estudio de una cohorte retrospectiva, y que el resultado de este estudio servirá para fijar una posible compensación económica si logra comprobarse que los sujetos que recibieron cierta exposición están en mayor riesgo de desarrollar la enfermedad en estudio. Si esta información fuese del conocimiento de los posibles participantes, pudiera ser que los sujetos que se saben expuestos y que desarrollaron la enfermedad en estudio participaran más frecuentemente que aquellos que desarrollaron la enfermedad y que se identifican como no expuestos. La participación diferencial antes mencionada se reflejaría en un déficit de eventos en el grupo de no expuestos, ya que los sujetos de este grupo que desarrollaron la enfermedad no tendrían la misma motivación para participar en el estudio y lo harían en menor proporción. Esta participación diferencial condicionaría una subestimación de la frecuencia del evento en el grupo no expuesto y una sobrestimación de la diferencia real.

Los estudios de casos y controles son particularmente susceptibles a este tipo de sesgo, ya que en la mayoría de sus aplicaciones se trata de estudios retrospectivos, es decir parten de sujetos de estudio que ya desarrollaron el evento de interés. Este tipo de sesgo se ha propuesto como una de las explicaciones a la asociación reportada entre el consumo de café y el cáncer de páncreas por McMahon y colaboradores.[1] En este estudio, los casos de cáncer de páncreas se seleccionaron de diferentes hospitales, mientras que los controles fueron seleccionados del mismo grupo médico de donde se había originado el diagnóstico de los casos. Al elegir este mecanismo de selección de controles, los investigadores incluyeron como controles a diferentes sujetos con patología gastrointestinal, dado que la mayoría de los casos de cáncer de páncreas habían sido notificados por el servicio de gastroenterología. Así, los controles representan un sector de la población con un consumo menos frecuente de café, lo que pudo haber condicionado una subestimación de la prevalencia de exposición en el grupo control y, por lo tanto, una asociación espuria entre el consumo de café y el cáncer de páncreas.

La detección diferencial es un tipo particular de sesgo de selección. Se origina cuando la prueba diagnóstica para detectar el evento se realiza con mayor frecuencia en el grupo expuesto. Un ejemplo de este sesgo se presentó en un estudio de casos y controles, en el que se observó una fuerte asociación entre el uso de estrógenos de reemplazo y el cáncer de endometrio.[2] Los casos en este estudio no repre-

sentaban adecuadamente los que se originaban en la población base, ya que las mujeres con síntomas sugerentes de cáncer de endometrio, que eran usuarias de estrógenos, eran hospitalizadas para diagnóstico con mayor frecuencia que las mujeres con los mismos síntomas, pero que no eran usuarias de estrógenos. Esta conducta de los médicos generó una sobre-selección de casos expuestos a estrógenos, lo que posiblemente ocasionó una asociación positiva falsa entre el uso de estrógenos y el riesgo de cáncer de endometrio en el estudio.

La falta de respuesta por parte de los participantes en un estudio puede también introducir sesgo de selección, siempre y cuando esté relacionada con la exposición o el evento en estudio; es decir, que la tasa de participación sea diferente para expuestos y no expuestos en los estudios de casos y controles, o para sujetos que desarrollaran el evento en el contexto de un estudio de cohortes. Sin embargo, no siempre la falta de respuesta se asocia con un sesgo de este tipo. Por ejemplo, si en un estudio de casos y controles, 10% de los casos elegidos y 30% de los controles rehúsan participar, no se incurrirá en sesgo de selección, a menos que la proporción de expuestos en los casos y en los controles que no aceptaron participar sea diferente a la del grupo respectivo. La falta de respuesta es un problema que tiene mayor relevancia para la generalización de los resultados a la población fuente.

Los estudios transversales, y de casos y controles que se basan en casos existentes (prevalentes), presentan importantes limitaciones relacionadas con los sesgos de selección, en particular cuando la enfermedad en estudio tiene una alta letalidad cercana al diagnóstico inicial, ya que los casos existentes tienden a sobre-representar a los sujetos con cursos más benignos de la enfermedad. Si el factor en estudio se asocia con la letalidad, la medida de efecto derivada de un estudio de prevalencia será sesgada, dado que los casos en estudio corresponden a sobrevivientes de la enfermedad. Esto dará como consecuencia que representen desproporcionadamente los casos no expuestos.



## SESGOS DE INFORMACIÓN

### Error de medición

Antes de abordar el concepto de sesgo de información explicaremos el significado de *error de medición*, pues éste supone la principal fuente de sesgos de información. No obstante, primero debemos señalar que en epidemiología no existen procedimientos libres de errores de medición, y que no todos los errores de medición son fuente de sesgos de información. En segundo término es conveniente recordar que los errores de medición pueden ser no diferenciales, cuando el grado de error del instrumento o técnica empleada es el mismo para los grupos que se comparan, y diferenciales, cuando el grado de error se distribuye de manera diferente en cada uno de los grupos estudiados.

Para comprender mejor la distinción entre los errores de medición diferenciales y no diferenciales, analizaremos un ejemplo. En un estudio hipotético de casos y controles para evaluar la asociación entre tabaquismo e infarto agudo del miocardio, los casos se identificaron en las salas de urgencias del Instituto Mexicano del Seguro Social (IMSS) en el momento de su ingreso, y los controles se seleccionaron de manera aleatoria entre los derechohabientes que son vecinos de cada caso. Supóngase que, en un primer escenario, la exposición al tabaco se evalúa determinando la presencia

de un marcador biológico de exposición que se mide en sangre. En este primer escenario, puede existir cierto grado de error en las determinaciones del biomarcador en sangre; sin embargo, es posible suponer que el error es similar para los dos grupos, por lo que se considera no diferencial. En contraste, si la exposición al tabaco se hubiera evaluado mediante un cuestionario, la calidad de la información dependería, en parte, de la memoria de los participantes. Si los casos, dado que sufrieron el evento, tuvieran un estímulo mayor para recordar o participar, entonces la calidad de la información sería mejor en este grupo que la que podría obtenerse en el grupo control, situación que podría introducir un error diferencial. En general, el impacto de este último tipo de error es difícil de predecir, ya que puede sobre o subestimar la asociación real, a diferencia del error no diferencial que, en general, tiende a subestimar las asociaciones reales.



En el cuadro I se ofrece una clasificación general de las diferentes fuentes de error de medición. Es importante hacer notar que, durante el diseño de un estudio, deben establecerse todas aquellas características que deberán medirse para responder una pregunta de investigación. Una de las dificultades con las que se topa el investigador en esta etapa se deriva de la brecha o diferencia que se establece entre la entidad conceptual de una variable y su definición empírica que es, finalmente, la forma en que será medida. Esta diferencia estriba en la imposibilidad, la mayor parte de las ocasiones, de medir directamente la característica de interés. Por ejemplo, la exposición involuntaria al humo del tabaco podría definirse teóricamente como "la exposición de las células del tejido pulmonar a agentes carcinogénicos que resulta de la inhalación del humo del tabaco proveniente de las emisiones de fumadores activos". En la práctica, esta definición teórica (entidad conceptual) tendría que medirse mediante la utilización de algún indicador cercano de dicha condición, tal como "tiempo que pasa el sujeto de estudio junto a personas fumadoras". Estas definiciones empíricas las utilizan y plasman los investigadores en los instrumentos de medición que se elaboran ex profeso para un estudio, y que pueden medirse por medio de un cuestionario o un biomarcador. En este último ejemplo, la exposición involuntaria también podría medirse mediante la determinación de cotinina (un metabolito de la nicotina) en la sangre de los sujetos en estudio. Mientras que las estimaciones de exposición realizadas mediante un cuestionario estarán sujetas a diferentes tipos de error, como serían la memoria de los participantes y la comprensión de las preguntas, en el caso del biomarcador, serán fuentes de error los procedimientos utilizados para obtener la muestra, el transporte y conservación de la misma, así como los diferentes procedimientos de laboratorio que necesitan utilizarse para su medición.

Para ilustrar el impacto que pueden tener los diferentes tipos de error de medición en la estimación de las medidas de efecto es necesario mencionar brevemente los conceptos

**Cuadro I Clasificación del error de medición según su origen**

a) El observador
b) Sistema de medición
c) Los sujetos de estudio
   Memoria
   Entrenamiento
   Fatiga
d) El instrumento
e) Errores en las variables *proxy*
f) El procesamiento de datos
   Errores de codificación
   Formulación errónea de modelos estadísticos
g) Errores que dependen del tiempo

de sensibilidad y especificidad de los instrumentos de medición que se utilizan. La sensibilidad de un cuestionario o prueba diagnóstica mide la habilidad del instrumento para clasificar correctamente a los sujetos que tienen verdaderamente la condición en estudio. Así, un cuestionario para identificar fumadores involuntarios, con una sensibilidad de 90%, podrá identificar correctamente a 90% de los sujetos que están expuestos al humo del tabaco, mientras que 10% serán clasificados erróneamente como no fumadores involuntarios. En contraste, la especificidad indica la habilidad del instrumento para clasificar correctamente a los sujetos que no tienen la condición del estudio. Para este ejemplo, una especificidad de 99% indicaría que se clasificaría adecuadamente a 99% de los sujetos, y que 1% sería clasificado como expuesto involuntariamente al tabaco, cuando en realidad no lo está. La sensibilidad y especificidad de nuestros instrumentos puede ser evaluada en estudios de validación cuando se cuenta con un estándar de oro. El cuadro II indica cómo podrían estimarse estos parámetros.

Los estudios epidemiológicos incurren en un error de medición no diferencial de la exposición, cuando ésta se realiza por igual en los grupos que se comparan; es decir que se comete el mismo grado de error de medición en el grupo que tiene el evento y en el grupo sin dicho evento. En esta circunstancia, siempre y cuando el error no sea de gran magnitud y se comparen sólo dos poblaciones, la medida de efecto estimada tiende a subestimar la asociación real en la población blanco; en otras palabras, el estimador tiende o se aproxima al valor de no asociación.

La dificultad para medir correctamente la exposición involuntaria al humo del tabaco puede servir para ilustrar este tipo de error. Dado que no es factible medir directamente la cantidad de humo que inhala un no fumador, podría utilizarse un monitor personal que registre en forma acumulada la concentración de nicotina ambiental a la que el sujeto estuvo expuesto durante el tiempo relevante para el evento en estudio. Supóngase que se lleva a cabo un estudio longitudinal para examinar la asociación entre enfermedad respiratoria y exposición involuntaria al humo del tabaco. En este caso, el error de medición no diferencial de la exposición estaría dado por la habilidad del sujeto de llevar consigo el monitor, el tipo de ropa que utilice y la conservación y trato que le dé al instrumento de medición. En este caso podría suponerse que el grado de error será el mismo para los sujetos



**Cuadro II** **Estimulación de sensibilidad y especificidad para una variable dicotómica**

| | *Clasificación de la exposición o de la enfermedad* | |
|---|---|---|
| Clasificación de la exposición o de la enfermedad por el instrumento de medición | Sí | No |
| Sí | *a* | *b* |
| No | *c* | *d* |

$$\text{Sensibilidad} = \frac{a}{a+c}; \quad \text{Especificidad} = \frac{d}{b+d}$$

con y sin exposición al humo de tabaco, y que el impacto de dicho error será el de subestimar cualquier asociación real que exista en la población blanco.

Otro ejemplo que permite ilustrar el contexto en el que ocurre el error de medición no diferencial es cuando la exposición en estudio no considera el periodo de riesgo relevante en la ocurrencia de una enfermedad. Si se estudia una enfermedad que tiene un periodo de latencia prolongado, como es el caso de la gran mayoría de las enfermedades crónicas, y el factor de exposición que se cree asociado con dicha enfermedad se mide en una sola ocasión, sin tomar en consideración la variabilidad de dicho factor durante el tiempo, se incurrirá en un error de medición no diferencial. De manera más explícita, puede decirse que, si se desea estimar la asociación entre el consumo de colesterol y enfermedad coronaria, para lo cual medimos la ingesta de colesterol en un solo día, no puede asegurarse que se haya medido correctamente la exposición, dado que la ingesta varía considerablemente a través del tiempo en un mismo sujeto y, por lo tanto, la medición correspondiente a un solo día no reflejaría adecuadamente la ingesta de colesterol durante el periodo de riesgo relevante.

A continuación se ofrece un ejemplo hipotético para ilustrar el efecto del error no diferencial. En el cuadro III se resumen los datos de un estudio de seguimiento por 12 años en hombres de 45-64 años (n= 7 221), en el que se relacionó la inactividad física con el riesgo de enfermedad coronaria. Los expuestos fueron quienes no hacían ejercicio físico o lo hacían moderadamente. Asumamos que la medida de efecto en este primer escenario se estima sin error aleatorio y corresponde a una razón de incidencia acumulada de 1.41.

Supongamos ahora que la determinación del evento se realiza con una prueba diagnóstica con una sensibilidad de 0.80 y especificidad de 1.0, y que la prueba se aplica de igual mane-

**Cuadro III** **Relación entre inactividad física y riesgo de enfermedad coronaria estimada sin error aleatorio**

| | Inactividad física | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 299 | 107 | 406 |
| No | 4 503 | 2 312 | 6 815 |
| Total | 4 802 | 2 419 | 7 221 |

Razón de incidencias acumuladas = (299/4 802) ÷ (107/2 419) = 1.41

**Cuadro III-A** **Relación entre inactividad física y riesgo de enfermedad coronaria, estimada con una prueba diagnóstica con sensibilidad de 0.80 y especificidad de 1.0**

| | Inactividad física | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 239 | 86 | 325 |
| No | 4 563 | 2 333 | 6 896 |
| Total | 4 802 | 2 419 | 7 221 |

Cálculos requeridos:

a) expuestos con el evento: $299 \times 0.80 \cong 239$
diferencia: $299 - 239 = 60$

b) no expuestos con el evento: $107 \times 0.80 \cong 86$
diferencia: $107 - 86 = 21$

c) expuestos sin el evento más los expuestos con el evento que fueron clasificados erróneamente como sanos: 4 503 + 60 = 4 563

d) no expuestos sin el evento más los no expuestos con el evento clasificados erróneamente como sanos: 2 312 + 21 = 2 333

Razón de incidencias acumuladas = (239/4 802) ÷ (86/2 419) = 1.39

**Cuadro III-B** **Relación entre inactividad física y riesgo de enfermedad coronaria, estimada con una prueba diagnóstica con sensibilidad de 0.80 y especificidad de 0.98**

| | Inactividad física | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 329 | 132 | 461 |
| No | 4 473 | 2 287 | 6 760 |
| Total | 4 802 | 2 419 | 7 221 |

Cálculos requeridos:

a) En la población expuesta:

299 × 0.80 = 239

diferencia: 299 – 239 = 60

4 503 × 0.98 = 4 413

diferencia: 4 503 – 4 413 = 90

Expuestos con el evento: 239 + 90 = 329

Expuestos sin el evento: 4 413 + 60 = 4 473

b) En la población no expuesta:

107 × 0.80 = 86

diferencia: 107 – 86 = 21

2 312 × 0.98 = 2 266

diferencia: 2 312 – 2 266 = 46

No expuestos con el evento: 86 + 46 = 132

No expuestos sin el evento: 2 266 + 21 = 2 287

Razón de incidencias acumuladas = (329/4 802) ÷ (132 / 2 419) = 1.26

ra en expuestos y no expuestos. Los resultados se presentan en el cuadro III-A, donde puede observarse que existe una pequeña subestimación del riesgo real, pues la razón de incidencias acumuladas disminuye de 1.41 a 1.39.

Si además suponemos una especificidad de 98%, entonces el error es de mayor impacto ya que la razón de incidencias acumuladas estimada bajo este supuesto sería de 1.26 (cuadro III-B).

### Sesgos de información

El sesgo de información se refiere específicamente a los errores en la estimación de la asociación que se introducen durante la medición de la exposición, los eventos u otras covariables, que se presentan de manera diferencial entre los grupos que se comparan y que ocasionan una conclusión errónea respecto de la hipótesis que se investiga. Una fuente de sesgo de información puede ser cualquier factor que influya de manera diferencial en la calidad de las mediciones que se realizan en los grupos expuesto y no expuesto (en el contexto de los estudios de cohorte) o entre los casos y controles (en el contexto de los estudios de casos y controles).

Como se mencionó, cuando el error de medición es diferencial se introduce un sesgo de información que, a diferencia del error aleatorio —que siempre provoca una subestimación— puede generar conclusiones erróneas en cualquier dirección y por lo tanto puede provocar tanto sobrestimación como subestimación de la asociación real. Varias situaciones que pueden asociarse con este sesgo. En la etapa de diseño del estudio puede favorecerse la introducción de sesgo si se decide recoger la información sobre la exposición después de que se han identificado los casos de la enfermedad de interés, y la presencia de la enfermedad modifica el informe de la exposición (ya sea exagerándola o minimizándola). En el caso del estudio de exposición involuntaria al humo del tabaco antes referido, si los investigadores utilizaran un cuestionario como instrumento para la evaluación de tipo retrospectivo no es difícil suponer que las madres cuyos hijos tuvieron problemas respiratorios tratarán de recordar con mayor detalle y exactitud la exposición al humo del tabaco, en comparación con aquellas cuyos hijos permanecieron sanos durante el periodo de estudio; o que las madres fumadoras cuyos hijos tuvieron enfermedad respiratoria informaran una menor expo-

sición como un mecanismo compensatorio de culpa. En ambas situaciones es claro que puede incurrirse en sesgos de información.

El sesgo de información de la exposición ocurre también con frecuencia durante la fase de campo. Un caso de sesgo de información del entrevistador puede ocurrir cuando el personal de campo conoce la condición de enfermo o no-enfermo del entrevistado, y además si tiene conocimiento de las hipótesis u objetivos del estudio. Ante este hecho, el entrevistador puede hacer más énfasis en aquellas preguntas que miden la exposición, o inducir las respuestas de los entrevistados mediante lenguaje corporal o la aportación de información adicional sobre dichas preguntas. En este sentido, es recomendable realizar un buen entrenamiento de los entrevistadores y mantenerlos ciegos tanto en lo que respecta a la hipótesis del estudio como a la condición de caso-control o de expuesto-no expuesto de los participantes del estudio.



El uso de informantes sustitutos (*proxy*) también puede dar lugar a un sesgo de información de la exposición. Como se mencionó, con cierta frecuencia los estudios epidemiológicos deben recurrir a informantes que no son los sujetos de estudio, ya sea por alguna incapacidad de éstos o porque ya fallecieron. Dependiendo del tipo de relación que tiene o tuvo el informante sustituto con el sujeto de estudio, las respuestas que éste proporcione sobre la exposición de interés pueden oscilar entre la negación o la exageración. Es probable que si la exposición se refiere a una característica considerada negativa, y el informante sustituto es el o la cónyuge del sujeto de estudio, las respuestas proporcionadas se modifiquen según la relación entre ambas personas.

Para ilustrar el impacto del sesgo de información, supongamos ahora que la relación descrita en el ejemplo anterior entre ejercicio y enfermedad coronaria se evalúa mediante un estudio de casos y controles. En este estudio se identificarían 406 casos y para cada caso se seleccionarían dos controles. Los resultados hipotéticos se presentan en el cuadro IV. La razón de momios (RM) estimada del estudio es de 1.41.

Supongamos que los casos hacen un mejor esfuerzo por recordar y reportar el patrón de ejercicio que realizaban, dado que sufrieron un evento coronario, y que la sensibilidad con que informan la exposición es de 95%, mientras que los controles reportan con una sensibilidad de 80%. Los resultados se presentan en el cuadro IV-A. En este escenario, el sesgo introducido por la mejor calidad del informe de actividad física de los casos condiciona una sobrestimación de la razón de momios.

Supongamos ahora que los casos tienden a subestimar el reporte de ejercicio y que la sensibilidad en este grupo es de 70%, mientras que en el grupo control el cuestionario utilizado clasifica adecuadamente la exposición en 90% de los sujetos. Bajo los supuestos de este escenario, el sesgo de reporte condicionaría una subestimación del efecto de no hacer ejercicio (cuadro IV-B). Como puede apreciarse, el efecto puede condicionar desviaciones hacia cualquier dirección, por lo que es fundamental para cualquier estudio evitar en la medida de lo posible introducir sesgos.

**Cuadro IV** **Relación entre inactividad física y riesgo de enfermedad coronaria, estimada sin error diferencial**

| | **Inactividad física** | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 299 | 107 | 406 |
| No | 540 | 272 | 812 |
| Total | 839 | 379 | 1 218 |

Razón de momios = 299 × 272 / 540 × 107 = 1.41

**Cuadro IV-A** **Relación entre inactividad física y riesgo de enfermedad coronaria, estimada a partir de un informe de la exposición con sensibilidad de 95% en los casos y 80% en los controles**

| | Inactividad física | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 285 | 121 | 406 |
| No | 432 | 380 | 812 |
| Total | 717 | 501 | 1 218 |

Cálculos requeridos:

a) En los casos:
   $299 \times 0.95 \cong 285$; diferencia: $299 - 285 = 14$
   $107 \times 1.0 = 107$; diferencia: $107 - 107 = 0$
   Casos expuestos: $285 + 0 = 285$
   Casos no expuestos: $107 + 14 = 121$

b) En los controles:
   $540 \times 0.80 \cong 432$; diferencia: $540 - 432 = 108$
   $272 \times 1.0 = 272$; diferencia: $272 - 272 = 0$
   Controles no expuestos: $432 + 0 = 432$
   Controles expuestos: $272 + 108 = 380$

Razón de momios = $285 \times 380 / 432 \times 121 = 2.07$

**Cuadro IV-B** **Relación entre inactividad física y riesgo de enfermedad coronaria, estimada a partir de un informe de la exposición con sensibilidad de 70% en los casos y 90% en los controles**

| | Inactividad física | | |
|---|---|---|---|
| *Evento coronario* | *Sí* | *No* | *Total* |
| Sí | 209 | 197 | 406 |
| No | 486 | 326 | 812 |
| Total | 695 | 523 | 1 218 |

Cálculos requeridos:

a) En los casos:
   $299 \times 0.70 \cong 209$; diferencia: $299 - 209 = 90$
   $107 \times 1.0 = 107$; diferencia: $107 - 107 = 0$
   Casos expuestos: $209 + 0 = 209$
   Casos no expuestos: $107 + 90 = 197$

b) En los controles:
   $540 \times 0.90 \cong 486$; diferencia: $540 - 486 = 54$
   $272 \times 1.0 = 272$; diferencia: $272 - 272 = 0$
   Controles expuestos: $486 + 0 = 486$
   Controles no expuestos: $272 + 54 = 326$

Razón de momios = 209 x 326 / 486 x 197 = 0.71



## SESGOS DE CONFUSIÓN

Todos los resultados derivados de estudios observacionales están potencialmente influidos por este tipo de sesgo. El sesgo de confusión puede resultar en una sobre o subestimación de la asociación real. Existe sesgo de confusión cuando observamos una asociación no causal entre la exposición y el evento en estudio, o cuando no observamos una asociación real entre la exposición y el evento en estudio por la acción de una tercera variable que no es controlada. Esta(s) variable(s) se denomina(n) factor(es) de confusión o confusor(es). Los resultados de un estudio estarán confundidos cuando los resultados obtenidos en la población en estudio apoyen una conclusión falsa o espuria sobre la hipótesis en evaluación, debido a la influencia de otras variables que no fueron controladas adecuadamente, ya sea durante la fase de diseño o de análisis. En este contexto, es fuente posible de sesgo de confusión cualquier variable asociada con la exposición que, además, esté causalmente asociada

con el evento en estudio y se encuentre distribuida de manera diferencial entre los grupos que se comparan, ya sea entre expuestos y no expuestos en el contexto de los estudios de cohorte, o entre casos y controles en el ámbito de los estudios de casos y controles.

En los estudios observacionales, el sesgo de confusión puede entenderse como un problema de comparabilidad, cuyo origen está ligado a la imposibilidad de realizar una asignación aleatoria de la exposición en los sujetos de estudio. El objetivo de la asignación aleatoria de los tratamientos (que en este caso es la exposición) en los estudios experimentales es intentar la formación de grupos homogéneos en lo que se refiere a todas las características que puedan influir en el riesgo de desarrollar el evento (edad, sexo, masa corporal u otras características que no puedan medirse). Lo que pretende lograrse es que los grupos sean similares en todo, excepto en la exposición a evaluar.



Analícese el siguiente ejemplo: cuando Doll y Hill publicaron sus primeros resultados derivados del seguimiento de cerca de 40 mil médicos, y reportaron que el cáncer de pulmón era considerablemente más frecuente en los fumadores,[3,4] las críticas no se hicieron esperar. Se argumentó que la asociación observada por estos investigadores entre el tabaquismo y el cáncer de pulmón se debía muy probablemente a la acción de una tercera variable, a una susceptibilidad genética que predisponía tanto al cáncer de pulmón como al gusto por el tabaco; con este argumento se descartaba el posible efecto carcinogénico del humo de tabaco. La crítica anterior implica que el gusto por fumar no se da al azar, sino que coexiste un factor genético, el cual es la verdadera causa de cáncer de pulmón. Para poder separar el efecto del cigarrillo del de la susceptibilidad genética, habría que realizar un experimento: podrían seleccionarse 40 mil individuos y asignarlos de manera aleatoria a dos tratamientos experimentales: fumadores (expuesto) y no fumadores (no expuesto). Al asignarlos de forma aleatoria puede suponerse que la distribución del factor genético que causa cáncer de pulmón en los grupos formados mediante este procedimiento aleatorio será la misma, es decir, si la aleatorización se realizó de manera correcta, el porcentaje de sujetos genéticamente susceptibles al cáncer de pulmón sería, en promedio, el mismo en ambos grupos, por lo que si observáramos diferencias en la ocurrencia de cáncer de pulmón entre los fumadores y no fumadores, ésta no podría atribuírsele al factor genético, pues éste afectaría de igual manera a ambos grupos. Sin embargo, es claro que este tipo de experimento sería imposible de realizar.

A diferencia de los estudios experimentales, en que los sujetos en estudio se asignan de forma aleatoria a los grupos experimentales y, por lo tanto, los posibles confusores quedan igualmente distribuidos entre los grupos en contraste, en los estudios observacionales los sujetos reciben la exposición por muy diferentes motivos que dependen, en gran medida, de patrones culturales y socioeconómicos. Entonces, ¿qué es una variable confusora o un confusor? Se trata de variables que están asociadas con el evento en estudio de manera causal, es decir, que son factores de riesgo para el evento en estudio. Además, esta asociación debe darse por un mecanismo causal diferente al que actúa en la exposición del estudio y debe ser independiente de la misma. Esto último significa que el confusor debe ser un factor de riesgo tanto para el grupo expuesto como para el no expuesto. Los sesgos de confusión en estudios de cohorte ocurren cuando los grupos expuesto y no expuesto no son comparables. Por ejemplo, si fuera de interés evaluar en un estudio de cohorte el efecto del consumo de café sobre el riesgo de infarto del miocardio y se establecieran dos grupos de contraste, uno conformado por bebedores

**Cuadro V** **Estrategias para el control de la confusión en los estudios epidemiológicos**

| *Fase* | *Estrategia* | *Efecto* |
|---|---|---|
| Diseño | Aleatorización | Permite que las variables se distribuyan similarmente en los grupos de estudio, y los hace comparables en todo, excepto en la variable de exposición. |
| | Restricción | Limita la participación en el estudio a sujetos que son similares, respecto a la variable de confusión. |
| | Pareamiento | Iguala en el proceso de selección a los grupos de comparación en relación con los factores de confusión. |
| Análisis | Estandarización | Permite comparar los grupos de estudio, cuando la distribución del confusor es la misma. |
| | Estratificación | Estima la medida de efecto en subgrupos que son similares en relación con los factores de confusión |
| | Modelos múltiples | Estima el efecto de la exposición, y mantiene constantes los valores del factor confusor |

de café y otro por no bebedores de café, tendría que considerarse el tabaquismo como un posible confusor, ya que está asociado con el consumo de café y es un factor de riesgo para infarto de miocardio. En este estudio hipotético, ignorar las diferencias en el consumo de tabaco podría hacernos concluir que existe una falsa asociación positiva entre el consumo de café y el riesgo de infarto del miocardio. En el contexto de los estudios de casos y controles, ocurre cuando hay diferencias entre casos y controles en relación con una tercera variable, lo que puede ocurrir simplemente por la existencia de confusión en la población fuente o ser generada por el mismo proceso de selección.

¿Qué puede hacerse para prevenir la ocurrencia de sesgos de confusión? En el cuadro V se listan los diferentes procedimientos tanto en la fase de diseño como en la de análisis, que nos permiten prevenir o corregir el efecto del sesgo de confusión, como puede deducirse de los diferentes procedimientos utilizados. Su común denominador es hacer los grupos de contraste lo más comparativamente posible en relación con variables externas. El objetivo es lograr que la única diferencia entre los grupos sea la característica en estudio.



## REFERENCIAS


1. MacMahon B, Yen S, Trichopoulos D *et al*. Coffee and cancer of the pancreas. N Engl J Med 1981; 304:630-633.
2. Feinstein HR, Horowitz RI. A critique of the statiscal evidence associating estrogens with endometrial cancer. Cancer Res 1978;38: 4001-4005.
3. Doll R, Hill AB. Mortality in relation to smoking: Ten years' observations of British doctors. BMJ 1964;1:1399-1410.
4. Doll R, Hill AB. Mortality in relation to smoking: Ten years' observations of British doctors (Concluded). BMJ 1964;1:1460-1467.

# Sesgos — Ejercicios

## REFERENCIA 1 (PREGUNTAS 1-4)

Lazcano–Ponce EC, Hernández–Ávila M, López–Carrillo L, Alonso-de–Ruiz P, Torres–Lobatón A, González-Lira G, Romieu I. Factores de riesgo reproductivo e historia de vida sexual asociados a cáncer cervical en México. Rev Invest Clin 1995;4:377-85.

El artículo reporta los resultados de una investigación realizada mediante un diseño de casos y controles con base poblacional, de acuerdo con los criterios establecidos en el capítulo de casos y controles para clasificarlos.

## REFERENCIA 2 (PREGUNTAS 5 Y 6)

Kershenobich D, Vargas F, García-Tsao G, Pérez Tamayo R, Gent M, Roffind M. Colchicine in the treatment of cirrhosis of the liver. N Eng J Med 1988; 318:1709-1713.

El artículo describe un ensayo clínico aleatorizado, diseñado para evaluar el efecto de la colchicina sobre la cirrosis hepática. El estudio se formuló de acuerdo con los criterios establecidos para este tipo de estudios en el capítulo sobre ensayos clínicos aleatorizados.

## REFERENCIA 3 (PREGUNTAS 7 Y 8)

Parra-Cabrera S, Hernández-Ávila M, Tamayo-y-Orozco J, López-Carrillo L, Meneses-González F. Exercise and reproductive factors as predictors of bone density among osteoporotic women in Mexico City. Calcif Tissue Int 1996;59:89-94.

El artículo corresponde a un estudio transversal —según los criterios señalados en el capítulo sobre estudios transversales— con el fin de evaluar la asociación entre ejercicio físico, algunos factores reproductivos y la densidad mineral ósea.



## PREGUNTAS

1. Bajo el supuesto de que existiera una relación inversa entre el tabaquismo y la sobrevida de cáncer cérvico-uterino, ¿cómo esperaría encontrar la medida de asociación si, en lugar de casos incidentes, los autores del primer estudio hubieran seleccionado casos prevalentes?
2. En el caso de que, efectivamente, hubiesen seleccionado sólo casos prevalentes, ¿en qué tipo de sesgo podrían incurrir?
3. Varios estudios epidemiológicos han encontrado una asociación positiva entre el cáncer cérvico-uterino y la pobreza. Si los autores del estudio únicamente hubieran incluido casos de hospitales privados, ¿cual habría sido el sesgo?; ¿qué dirección tendría su estimador?
4. Los autores encontraron que conforme se incrementaba el número de parejas sexuales se incrementaban las posibilidades de tener cáncer cervical. En los estudios de casos y controles, los casos suelen hacer un mayor esfuerzo por recordar los posibles factores de exposición, ¿qué tipo de sesgo se comete si no se controla este fenómeno?, ¿qué efecto tendría sobre el estimador?

5. Los autores informan que realizaron una asignación aleatoria de los sujetos tanto en el grupo de intervención como en el grupo control. Ni los médicos ni los pacientes sabían a qué grupo pertenecían estos últimos. La asignación aleatoria y el doble cegamiento son estrategias utilizadas para prevenir un sesgo entre:
   a) los investigadores
   b) el grupo experimental
   c) el grupo control
   d) todos los participantes
6. Durante el estudio, 10 pacientes del grupo de intervención y 9 del grupo control se perdieron durante el seguimiento. Si 9 de los 10 pacientes del grupo de intervención hubieran abandonado el estudio debido a la presencia de diarrea como consecuencia del uso de colchicina, ¿podría haberse introducido un sesgo de selección por pérdidas en el seguimiento?, ¿cómo esperaría encontrar la sobrevida del grupo de intervención, si los sujetos con diarrea tuvieran peor pronóstico que los que no la tuvieron?
7. ¿Qué tipo de diseño le sugiere el título del artículo?
8. Los autores encontraron una asociación positiva entre el ejercicio moderado y la densidad mineral ósea de la región femoral. Si quisiéramos imaginar una posible relación causal, ¿cuál sería la limitación más importante de este estudio, considerando que se trata de un diseño transversal?

## RESPUESTAS

1. La razón de momios se habría subestimado ya que, muy probablemente, una gran parte de los casos con antecedente de tabaquismo hubieran muerto. En otras palabras, los casos que sobrevivieron, con seguridad, tendrían un antecedente de tabaquismo muy bajo en relación con los que murieron.
2. Un sesgo de selección.
3. Estarían cometiendo un sesgo de selección, debido a que la selección de los casos estaría relacionada con un nivel socioeconómico alto. Lo anterior tendría como consecuencia una subestimación del efecto real, y la dirección del estimador podría invertirse.
4. Se trata de un sesgo de información, conocido en la bibliografía como sesgo de recordatorio. El estimador estaría sobrestimando la asociación real.
5. El doble cegamiento permite prevenir sesgos originados en el comportamiento de cualquiera (es decir, de todos) de los participantes durante el estudio.
6. Se habría introducido un sesgo de selección por pérdidas en el seguimiento, si la diarrea estuviera asociada con el pronóstico de la enfermedad. La sobrevida en el grupo de intervención sería mayor, debido a que la mayoría de los pacientes con diarrea que se perdieron durante el estudio —y de los que se desconoce su suerte— probablemente habrían muerto durante el seguimiento.
7. El título del artículo sugiere un estudio prospectivo, ya que sugiere que el ejercicio y los factores reproductivos son predictores de la densidad mineral ósea. Sin embargo, en el contexto del modelaje estadístico en epidemiología, en muchas ocasiones se emplea el término "variable predictora" como sinónimo de "variable independiente", sin implicar temporalidad.

8. La principal limitación de este trabajo, que es inherente a todos los estudios transversales, es la ausencia de dirección de la temporalidad entre los factores de exposición y el evento de interés. Con este diseño es imposible determinar si el ejercicio es un determinante de la densidad mineral ósea o viceversa.

# XI Introducción al análisis estadístico

La aplicación de la estadística en el análisis epidemiológico es una estrategia fundamental para cuantificar las asociaciones presentes entre las variables que intervienen en los fenómenos de salud. Esta cuantificación permite entender las relaciones subyacentes en dichos fenómenos, evaluar la precisión y predecir su curso.

Debido a la naturaleza tan diversa de la información que se obtiene en un estudio epidemiológico, los procedimientos estadísticos deben permitirnos, en primer lugar, hacer un análisis exploratorio de la información, con el fin de detectar posibles errores de captura o la presencia de valores atípicos, además de conocer con detalle la población de estudio. Esta primera fase también es útil para describir algunas asociaciones simples que más tarde facilitarán un análisis más elaborado. La etapa de análisis exploratorio de la información, fundamental en todo análisis estadístico, no está incluida en esta sección. Los autores suponen que el lector está familiarizado con las diversas técnicas estadísticas para la exploración de datos, mismas que, por lo general, son cubiertas en los cursos básicos de estadística. En esta sección, en cambio, se exponen con relativo detalle algunos de los procedimientos habitualmente usados en el análisis estadístico: la mejor manera para resumir, analizar y utilizar la información recolectada para contestar una pregunta de investigación fraguada en el ámbito de la epidemiología.

El resumen de la información puede realizarse por medio de la generación de un modelo estadístico, cuyo propósito es incorporar todas las piezas de información disponibles que son relevantes para el fenómeno de estudio.

Los cinco apartados de esta sección introducen al lector en el manejo de algunos de los modelos estadísticos más importantes utilizados en la investigación epidemiológica. No se pretende, de ningún modo, hacer una presentación exhaustiva de las diferentes técnicas de modelaje existentes, sino proporcionar al lector herramientas capaces de iniciarle en el análisis riguroso de su información estadística. Por esta razón se ha decidido presentar cada sección mediante el uso de ejemplos, con la finalidad de que el lector descubra paulatinamente los principales detalles técnicos y formas de aplicación de los modelos presentados. Queremos destacar que para un conocimiento más profundo de los modelos es recomendable la consulta de textos especializados en cada uno de los temas que aquí se presentan. Las referencias al final de cada capítulo pueden auxiliar en este propósito.

Las variables estudiadas deben corresponder al marco conceptual del área de investigación en cuestión: la epidemiología, la medicina, biología, economía y en general, de las disciplinas involucradas.

Los modelos estadísticos pretenden representar de la mejor manera la realidad que subyace detrás de las asociaciones encontradas y conocer su magnitud cuando se ponen en juego todas las variables estudiadas, o la mayoría. De ahí la importancia de generar "buenos" modelos, sus-

tentados desde punto de vista estadístico, pero epidemiológicamente útiles para responder adecuadamente a las preguntas de investigación. Sabemos que un modelo nunca será una representación exacta de los datos, que no podrán incluirse en él todas las observaciones necesarias y que nunca se ajustará perfectamente al total de la población de estudio. No obstante, debemos aspirar a que describa con el menor error posible las principales tendencias del fenómeno estudiado; identifique con elevada probabilidad las relaciones realmente existentes, y, en la medida de lo posible, descarte aquellas que son espurias. Ya que en el modelo estadístico elegido descansará gran parte de la respuesta a nuestro problema de estudio, es de vital importancia que reflexionemos en su construcción tanto como en su evaluación. El diseño con el que se obtuvo la información debe considerarse en la generación del modelo.

La valoración de qué tan bien puede el modelo estadístico representar un fenómeno epidemiológico depende, en gran medida, del conocimiento que tengamos sobre las relaciones que se dan entre las variables involucradas en el estudio. Esto implica que la elección del modelo no constituye, en absoluto, una selección arbitraria. Esta selección debe realizarse de acuerdo con los objetivos que persigue el estudio y el tipo de información disponible. Al referirnos al tipo de información debemos considerar, sobre todo, la escala de las variables del estudio y la escala de medición de la respuesta.

La sección presenta los modelos estadísticos de uso más frecuente en los diseños epidemiológicos, y hace especial énfasis en tres aspectos que consideramos fundamentales: 1) la comprensión de las características de la información que requiere la aplicación de cada modelo; 2) la interpretación de los resultados del análisis en el marco de la investigación epidemiológica y, finalmente, 3) la evaluación detallada del modelo propuesto.



La sección inicia con una presentación del modelo de regresión lineal, que es, quizás, el modelo más conocido en la estadística: modela la asociación entre una variable respuesta registrada como variable continua y un conjunto de variables independientes sobre las que no existe ninguna restricción en cuanto a la escala de medición utilizada. Dada la relevancia de este modelo, se le dedican dos capítulos para una mejor presentación. En el primero de ellos se describe con detalle el modelo de regresión lineal simple, para el cual la geometría en el plano cartesiano proporciona un instrumento gráfico de gran apoyo en la explicación de los conceptos fundamentales. En el segundo, se presenta la extensión del modelo simple al múltiple. La comprensión de los modelos que se presentan posteriormente descansa en gran parte en el entendimiento del modelo lineal.

Enseguida se presenta el modelo de regresión logística conocido como *logit.* En muchos estudios epidemiológicos el objetivo primordial consiste en explicar la asociación entre un grupo de factores de riesgo y la presencia o ausencia de una enfermedad (o cualquier otra condición de salud) en una población de estudio. La diferencia fundamental con el modelo lineal es que, en este caso, la variable dependiente ya no es continua, sino que refleja la presencia o ausencia de una condición de salud y comúnmente se codifica por medio de una variable, cuyo registro es *uno* si el evento está presente y *cero* si no lo está. Este método se utiliza frecuentemente en los estudios de casos y controles y en los transversales, donde la variable respuesta cumple con estas características. No obstante, las diferencias metodológicas entre estos dos tipos de diseño, especialmente en lo que se refiere a las estrategias de muestreo, obligan a realizar consideraciones especiales para el uso del modelo en cada uno de ambos diseños. Esta reflexión se aborda al final del capítulo.

El tercero y último de los modelos que se presentan es el modelo de riesgos proporcionales o modelo de Cox, muy importante dentro del área estadística conocida como análisis de supervivencia. Dadas las características peculiares que presenta la información recabada en los estudios en esta área, no solamente se presenta el modelo de riesgos proporcionales, sino que se hace una breve descripción de varias técnicas estadísticas que consideran estas peculiaridades y cubren diversas etapas en el proceso de análisis de la información que generan estos estudios.

La sección finaliza con un capítulo que permite clarificar el concepto de interacción, tan importante en epidemiología, pero muchas veces ignorado en la construcción de modelos multivariados. En esa sección se presentan conceptos fundamentales sobre interacción y modificación de efecto, y ejemplos en donde se aplica este concepto en el contexto de los modelos de regresión lineal y logística. Los ejemplos presentados corresponden a interacciones entre variables continuas y discretas, pues las interacciones con variables categóricas son, en realidad, casos especiales de las primeras, y la manera de evaluarlas e interpretarlas puede derivarse fácilmente de los procedimientos presentados en la primera parte del mismo apartado.

# XII Regresión lineal simple

*Daniela Sotres Álvarez*
*Martha María Téllez Rojo Solís*

## INTRODUCCIÓN

Una de las actividades más importantes de la epidemiología es la evaluación de los efectos de una exposición determinada en el desarrollo de una cierta condición de salud. Esta exposición puede ser, o bien potencialmente perjudicial (consumo de tabaco, exposición a ozono, ingestión de plomo, etc.), o bien potencialmente benéfica (uso de un programa de asistencia social, aplicación de una prueba de Papanicolaou, consumo de un fármaco, etc.). Mientras que para el primer grupo de exposiciones se esperan efectos nocivos a la salud, en el segundo grupo se esperan efectos positivos. Desde el punto de vista estadístico, lo común es que la exposición se asocie (idealmente de manera causal) con la variable de salud. La variable de salud es entonces "dependiente" de la exposición; y esta última, es decir, la exposición, es la "variable independiente". En la literatura, a la primera también se le conoce como variable de respuesta y a la segunda, como variable predictora o explicativa. Estos términos se usarán de forma indistinta a lo largo de los siguientes capítulos. La estadística se encarga de la evaluación de esta asociación desde un enfoque cuantitativo, y la metodología adecuada para el análisis respectivo dependerá de la escala de medición de cada una de estas variables. La interpretación apropiada de estas asociaciones dependerá fuertemente del diseño de estudio (cfr. capítulos sobre principales diseños de estudio).

La presente exposición iniciará por el caso más simple. Cuando se quiere saber si la media de una determinada condición que se puede medir en forma de variable continua es diferente en dos poblaciones se tienen diversas alternativas. Se puede usar, por ejemplo, una prueba paramétrica, como la prueba *t* de Student,[1] o una prueba no paramétrica, como la de Mann-Whitney.[1] En ambos casos, las poblaciones de estudio están clasificadas en dos grupos, y pueden identificarse por medio de una variable que tome sólo dos valores, donde cada uno de ellos indique la pertenencia a un grupo específico; es decir, se presenta una variable independiente dicotómica, binaria o indicadora. La variable de interés (o dependiente) es continua. En otro ejemplo, quiere compararse la capacidad pulmonar medida en litros (variable dependiente), entre niños expuestos a humo de tabaco y niños no expuestos a este contaminante (variable independiente), y en otro desea conocerse si la ingesta de un complemento alimenticio durante el embarazo (variable independiente) tiende a incrementar el peso al nacer (variable dependiente) de hijos de madres desnutridas al compararlo con un placebo. En todos los ejemplos anteriores, las variables dependientes son continuas y las variables independientes son dicotómicas.

Sin embargo, estas pruebas estadísticas no nos permiten abordar situaciones más complejas, tales como: 1) aquéllas donde se involucran variables de exposición con más de dos categorías o incluso, variables de exposición continuas; 2) aquéllas en las que es necesario ajustar la compa-

ración por posibles confusores, y 3) una mezcla de las dos anteriores. Una forma convencional de abordar la situación donde la variable de exposición tiene más de dos categorías, podría resolverse típicamente por medio de un análisis de varianza de una vía[1,2] o mediante alternativas no paramétricas como la prueba de Kruskal-Wallis,[1] entre otras. Sin embargo, los modelos de regresión lineal también permiten resolver este problema, ya que son una generalización de la prueba *t* y del análisis de varianza de una vía. Además, un uso muy común de estos modelos es que permiten comparar estadísticamente una variable continua entre dos o más poblaciones, pero por medio del ajuste de posibles confusores (a este análisis se le conoce como análisis de covarianza o ANCOVA). Una característica común entre la prueba *t*, el análisis de varianza de una vía, el ANCOVA y un modelo de regresión lineal, es que la variable dependiente es necesariamente continua y, al igual que en las dos primeras pruebas estadísticas, hay un supuesto de normalidad subyacente, el cual se abordará con más detalle en las siguientes secciones.

El análisis de regresión es un método estadístico que evalúa la asociación de una variable de respuesta $Y$ con una o más variables explicativas $X_1, X_2, \ldots, X_k$ ($k \geq 1$). En particular, cuando se habla de regresión lineal, se refiere a un tipo específico de modelo, el cual se propone como una descripción adecuada de la asociación postulada entre las variables: una línea recta para el caso en que se tenga una sola variable explicativa continua y un plano o hiperplano cuando se tengan dos o más variables explicativas. La pendiente de la recta en el caso particular de una variable independiente continua, y los coeficientes asociados con cada una para el caso de dos o más variables independientes, permitirán cuantificar la asociación entre la variable correspondiente y la de respuesta. Una recta muy inclinada (pendiente grande en valor absoluto) sugerirá que cambios pequeños en la variable independiente se asocian con grandes cambios en la media de la variable de respuesta y una recta casi horizontal (pendiente pequeña) será evidencia de una asociación débil en la que se requiere de grandes cambios en la variable independiente para lograr un cambio pequeño en la media de la variable de respuesta. Así, los parámetros de interés a estimar serán: la pendiente, cuando haya una sola variable independiente continua,* y los coeficientes asociados a cada una, cuando haya dos o más independientes (que, en lo sucesivo, se denominarán coeficientes de regresión).



Una pregunta pertinente sería: ¿por qué generar un modelo de regresión lineal simple si el coeficiente de correlación de Pearson también evalúa la asociación entre dos variables continuas? La razón es que el modelo de regresión lineal, además de determinar el grado y la dirección de la asociación, permite estimar el cambio promedio en unidades de la variable de respuesta $Y$ por un cambio en la variable explicativa $X$.

En este capítulo y el siguiente se estudiarán el modelo de regresión lineal, la estimación de los parámetros de interés y su interpretación. Este capítulo abordará el caso particular del modelo de regresión que sólo tiene una variable explicativa (es decir, $k = 1$) y que se conoce como modelo de regresión lineal simple; el capítulo siguiente se abocará al caso en que hay más de una variable explicativa ($k \geq 2$) y que se conoce como modelo de regresión lineal múltiple.

* Además, es necesario que la relación entre la variable independiente y la de respuesta sea lineal, porque de no serlo, podría ajustarse un polinomio y entonces no se estaría hablando de una recta.

## EL MODELO DE REGRESIÓN LINEAL SIMPLE

El siguiente ejemplo abordará los conceptos y notación requeridas. Entre 1998 y el año 2000, se realizó un estudio transversal en 434 mujeres posmenopáusicas trabajadoras del Instituto Mexicano del Seguro Social (IMSS), delegación Morelos, con el objetivo de estudiar algunos factores asociados con la densidad mineral ósea (`dmo`). Con esta idea en mente, se deseaba conocer la asociación entre la edad de la mujer (variable independiente o exposición) y la densidad mineral ósea (variable dependiente o de respuesta). Esta asociación se resume en la figura 1.

En general, el problema de regresión lineal simple consiste en estimar la "mejor" recta que se ajusta a una muestra aleatoria de $n$ pares de datos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Cada pareja corresponde a la medición de dos variables en un mismo sujeto, una variable de respuesta y una predictora. Un primer paso exploratorio indispensable, antes de generar un modelo de regresión lineal simple, consiste en graficar la variable independiente contra la dependiente para visualizar la relación que existe entre ambas. Si gráficamente vemos una relación aproximadamente lineal, entonces puede proponerse un modelo de regresión lineal simple que resuma la información de todos los puntos. El hecho de que el modelo propuesto sea precisamente una recta, implica restricciones importantes que conviene reflexionar desde el marco conceptual del estudio: el cambio estimado en la variable de respuesta $Y$ será el mismo para cualquier cambio de una unidad en la variable predictora $X$, independientemente de cuál sea el valor específico de $X$. Es importante señalar que la relación lineal se restringe a los valores de X que pertenecen a la región de exploración; es decir entre la mínima y la máxima X observadas.

En el ejemplo, el objetivo es determinar y cuantificar la asociación lineal entre densidad mineral ósea y edad, en mujeres posmenopáusicas. Como puede observarse en la figura 2, al aumentar la edad parece que disminuye la densidad mineral ósea y que esta relación es aproximadamente lineal. Una asociación aproximadamente lineal significa que se espera que un incremento de un año de edad en mujeres de 40 años tendrá el mismo impacto que un incremento de un año de edad en mujeres de setenta. Sin embargo, no se puede decir nada respecto al incremento de un año en mujeres de 17 años, porque no se estudiaron mujeres con edad cercana a los 17 años. La edad mínima fue de 41 años.

El modelo teórico utilizado para establecer la asociación lineal entre la variable de respuesta $Y$ y la variable explicativa $X$ es:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (1)$$

donde $\beta_0$ y $\beta_1$ son la ordenada al origen y la pendiente respectivamente, y $\varepsilon$ es el término de error considerado aleatorio. Este término de error incorpora toda la información de las variables ob-



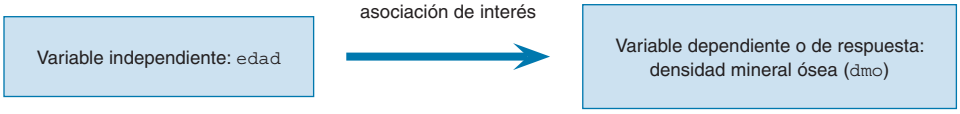



**Figura 1** **Asociación entre la edad de la mujer y la densidad mineral ósea**

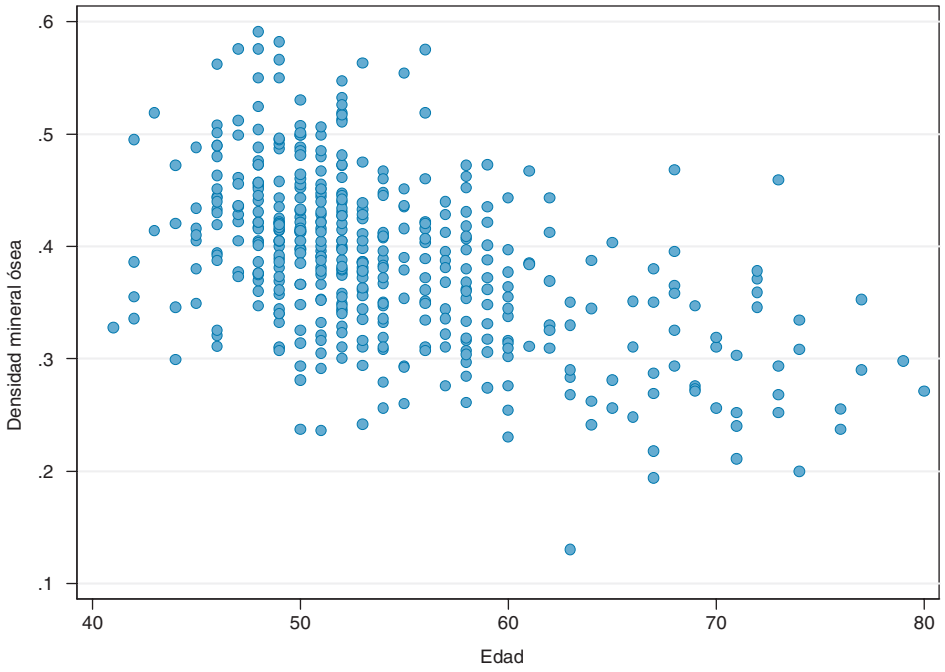
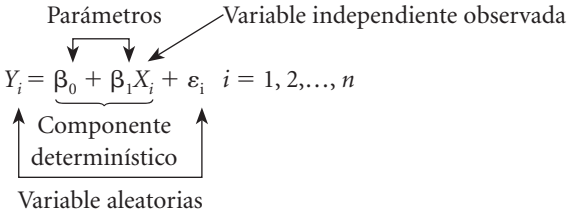





**Figura 2** **Gráfica de dispersión entre edad y densidad mineral ósea**

servables y no observables que podrían estar asociadas con la variable de respuesta (en nuestro ejemplo, `dmo`), que no está explicado por la variable independiente (`edad`). La aleatoriedad del término de error hace que la variable *Y* también sea una variable aleatoria; de ahí que, para cada valor fijo de *X*, haya una función de distribución de *Y*. En el siguiente esquema se muestra cada componente del modelo de regresión lineal simple para el *i*-ésimo sujeto.




Al suponer que la media de los errores es cero, el valor esperado de *Y*, dado un valor fijo de *X*, está dado por:

$$\mu_{Y|X} = E[Y \mid X] = \beta_0 + \beta_1 X$$

es decir, puede modelarse la media de $Y$ dado un valor fijo de $X$ [en notación matemática $E(Y|X)$], por medio de una función lineal de los parámetros $\beta_0$ y $\beta_1$. Así, el componente determinístico del modelo es la media de la variable de respuesta $Y$ en la población de mujeres con $X_i$ de edad. A continuación se describen los cuatro supuestos que subyacen en el modelo (1), y que son necesarios para poder hacer inferencias estadísticas[1] a partir de la muestra:

1. *Independencia de los errores*: los errores, entre cualesquiera dos observaciones diferentes, $\varepsilon_i$ y $\varepsilon_j$ son estadísticamente independientes y, como consecuencia, los valores de $Y$ dados los valores específicos de $X$ también lo son.
2. *Linealidad*: las medias de $Y$, dados los valores fijos de $X$, forman una línea recta.
3. *Homoscedasticidad* (homogeneidad de la varianza): la variabilidad del error es constante y es igual para todos los errores $\varepsilon_i$. Como consecuencia, para diferentes valores fijos de $X$, la varianza de $Y$ es la misma.
4. *Normalidad*: los errores tienen una distribución normal con media cero y varianza constante $\sigma^2$.

A partir del modelo teórico formulado y del supuesto distribucional de los errores se puede demostrar que:

$$Y \mid X \sim N(\mu_{Y|X}, \sigma^2)$$

En la figura 3 se ejemplifica la distribución de $Y$ para tres valores específicos de la variable independiente $X$, con el objetivo de ilustrar los supuestos de normalidad, homoscedasticidad y linealidad. Cada "campana" está modelando la distribución de $Y$ para una población particular (por ejemplo, para la edad específica $x_1$) y el supuesto es que la distribución de $Y$ es normal. Más aún, la varianza de $Y$ para diferentes poblaciones (por ejemplo, para las edades $x_1$ y $x_2$) es la misma. El supuesto de linealidad se representa con la recta que une las medias de $Y$. Generalmente, el supuesto de independencia se puede discutir en el contexto del diseño del estudio por medio de preguntas como: ¿las observaciones que generaron la muestra son independientes? En el ejemplo, esto equivaldría a hacerse la siguiente pregunta: ¿existen hermanas en el estudio de densidad mineral ósea, cuyo componente genético común pueda influir en su densidad, independientemente de los factores ambientales bajo estudio?

Estos supuestos son "requisitos" que tienen que cumplirse para que el modelo de regresión lineal se considere apropiado para el estudio particular. Cuando alguno de ellos no se cumple, es necesario recurrir a otros métodos estadísticos que no los requieren.[3]

Al utilizar la información muestral para estimar este modelo, se dice que se tiene un modelo estimado y se especifica como:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \qquad (2)$$

donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores de la ordenada al origen y la pendiente respectivamente. Se denotará con el símbolo ^ (en estadística, se le conoce como *gorro*) a los estimadores calculados a partir de la muestra. Por convención, en estadística se utilizan las letras mayúsculas para denotar el nombre de las variables y las minúsculas para los valores observados que toman estas va-

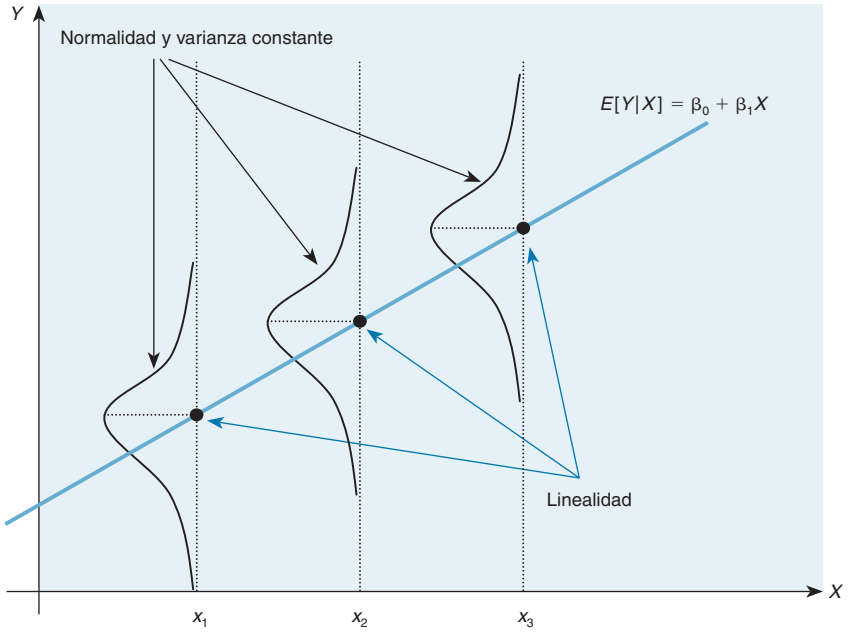





**Figura 3** **Representación del supuesto de normalidad, homoscedasticidad y linealidad, en un modelo de regresión lineal**

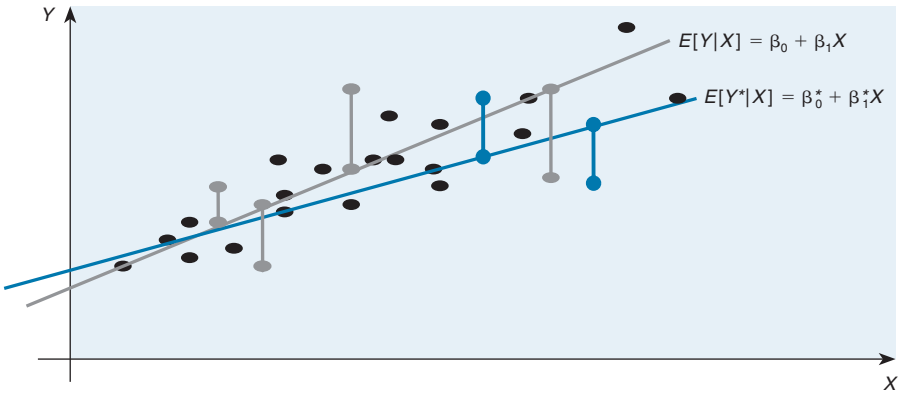



**Figura 4** **Desviaciones de los valores observados a dos rectas candidatas de regresión lineal simple**

riables. Así, *Y* (mayúscula) denota la variable dependiente, mientras que *y* (minúscula), un valor observado particular.

En la figura 4 puede verse un conjunto de datos y dos rectas que modelan la relación lineal entre *X* y *Y*; pero, ¿qué criterio puede adoptarse para decidir cuál de las dos rectas es mejor? Esta pregunta sería equivalente a: ¿cómo se encuentran los mejores estimadores de los parámetros desconocidos $\beta_0$ y $\beta_1$ de la recta dada por la ecuación (1)? Existen diversos métodos para estimar $\beta_0$ y $\beta_1$. En las siguientes secciones, se discuten dos métodos de estimación: el de mínimos cuadrados y el de máxima verosimilitud.

## ESTIMACIÓN DE LOS PARÁMETROS DE REGRESIÓN LINEAL SIMPLE POR MÍNIMOS CUADRADOS

El método más intuitivo (que no requiere de ningún supuesto sobre la forma de la distribución del error) es el método de mínimos cuadrados (*i.e.* no se requiere la normalidad del supuesto 4). Un primer acercamiento para escoger la recta que mejor se ajusta a los datos consiste en cuantificar las diferencias entre los valores observados y los valores predichos por todas las posibles rectas; la mejor será aquélla que minimice simultáneamente estas diferencias. Una posibilidad muy intuitiva sería minimizar la suma de todas las diferencias, pero esto tendría el inconveniente de que algunas serían positivas y otras negativas, dependiendo de si el valor observado queda por arriba o por debajo de la recta respectiva, por lo que se cancelarían unas con otras y se daría lugar a una idea falsa de la magnitud total de las desviaciones. Otra opción sería minimizar la suma de los valores absolutos, pero esto es matemáticamente difícil de lograr. Entonces, una mejor estrategia que evade los problemas de las propuestas anteriores es minimizar la suma de los cuadrados de estas diferencias. De ahí el nombre de mínimos cuadrados. Entonces, para obtener los estimadores de la ordenada al origen y la pendiente, hay que encontrar los valores únicos $\hat{\beta}_0$ y $\hat{\beta}_1$ que hacen mínima la siguiente función:

$$f(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left[ y_i - (\beta_0 + \beta_1 x_i) \right]^2$$

Al realizar el proceso de minimización matemática (el cual no se presenta en este texto pero que puede consultarse en textos especializados),[2,3] los estimadores que se obtienen son:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

donde $\bar{x}$ y $\bar{y}$ denotan los promedios muestrales de las variables *X* y *Y*, respectivamente. A los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ se les denomina comúnmente *coeficientes de regresión estimados.* A la diferencia entre el valor observado y el valor predicho $y_i - \hat{y}\ i = 1, 2\ldots, n$ se le conoce como residuo; posteriormente se verá que estos residuos serán de gran utilidad para evaluar si los supuestos del modelo se cumplen. La figura 5 muestra que, de las dos rectas presentadas en la figura 4, la

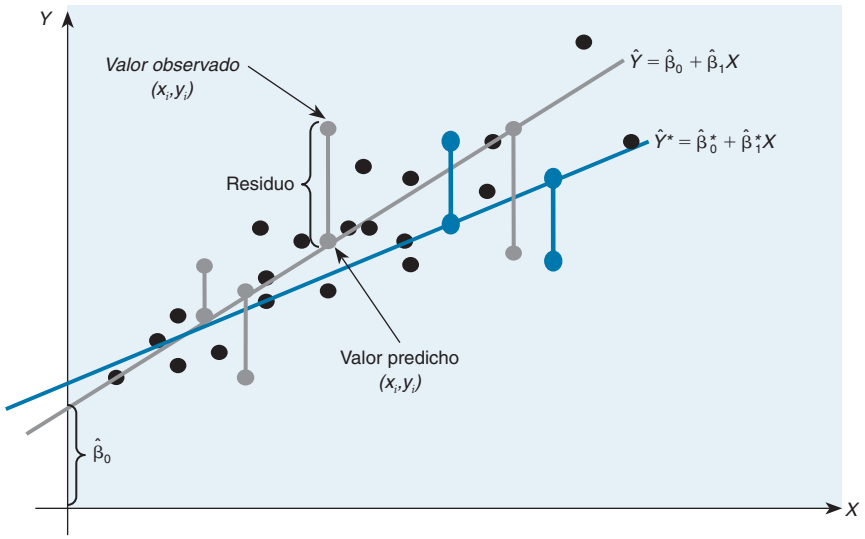



**Figura 5** **La recta de regresión lineal simple**



mejor es aquélla que se obtuvo por el método de mínimos cuadrados y que se presenta en color azul. Para resaltar que ésta es "la mejor recta" que puede obtenerse, se le llamará $\hat{Y} = \hat{\beta}_0$ y $\hat{\beta}_1 X$. En esta figura también se ilustra la ordenada al origen, el valor observado $y_i$, el valor predicho $\hat{y}_i$ y el residuo, correspondientes a la $i$-ésima observación.

Ahora bien, cualquier paquete estadístico calcula los estimadores de los coeficientes de regresión al especificar el modelo de regresión lineal. En STATA, el comando para hacer una regresión lineal es **`regress`**, que va seguido de la variable dependiente y después de las variables independientes. En el caso de regresión lineal simple, sólo hay una variable independiente, por lo que la sintaxis es: **`regress, YX.`**

Para el ejemplo, el modelo teórico que se propone es:

$$\text{dmo} = \beta_0 + \beta_1 \times \text{edad} + \varepsilon$$

Para poder estimar los parámetros $\beta_0$ y $\beta_1$, se utiliza el comando y la salida de STATA que se presentan a continuación. La salida de STATA se dividirá en tres partes, las cuales se estudiarán a lo largo de este capítulo:

A. cuadro de análisis de la varianza;
B. estadísticas y prueba global *F*;
C. coeficientes de regresión, errores estándar e intervalos de confianza, así como las pruebas de hipótesis para determinar si los coeficientes de regresión son estadísticamente diferentes de cero.

A

```
regress  dmo edad
```

```
      Source |       SS       df       MS
-------------+------------------------------
       Model |  .621464026     1  .621464026
    Residual |  1.82188781   432  .004217333
-------------+------------------------------
       Total |  2.44335184   433  .005642845
```

B

```
Number of obs =     434
F(  1,   432) =  147.36
Prob > F      =  0.0000
R-squared     =  0.2543
Adj R-squared =  0.2526
Root MSE      =  .06494
```

C

```
------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        edad |  -.0052193     .00043   -12.14   0.000    -.0060644   -.0043742
       _cons |   .6710929   .0234957    28.56   0.000     .6249129     .717273
------------------------------------------------------------------------------
```

## INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESIÓN

Por el momento, se determinará el modelo estimado (también conocido como modelo ajustado) e interpretarán los estimadores de los coeficientes de regresión a partir de la parte C de la salida de STATA presentada. En el caso de la regresión lineal simple, hay tres parámetros a estimar: $\beta_1$ que es el coeficiente de regresión asociado con la variable independiente (`edad`), la ordenada al origen $\beta_0$ y la varianza del error, $\sigma^2$. El estimador del coeficiente de regresión asociado con la variable `edad` es -0.0052. El signo negativo muestra que la relación entre `edad` y `dmo` es inversa: al comparar mujeres de diferentes edades, las de mayor `edad` tienen en promedio, menor `dmo`. Esto confirma lo observado en la figura 2; sin embargo, lo más interesante es que se puede cuantificar esta asociación. La `dmo` es, en promedio, 0.0052 g/cm$^2$ menor, por cada incremento de un año de edad. Es muy importante resaltar, primero, que se comparan dos mujeres distintas y no a la misma mujer al incrementar un año su edad (esto último se haría con un modelo de regresión para datos longitudinales que no se cubre en este texto). Segundo, la diferencia estimada es un promedio; es decir, en promedio, mujeres de mayor `edad`, tendrían mediciones de `dmo` menores. También puede determinarse la diferencia promedio en la `dmo` entre edades que difieren cinco o diez años, si se multiplica el coeficiente de regresión por la diferencia de edades que interesa; es decir, un incremento en la `edad` de cinco o diez años implica una disminución promedio en la `dmo` de 0.026 g/cm$^2$ o 0.052 g/cm$^2$, respectivamente. Nótese también que el coeficiente de regresión asociado a la variable `edad` depende de la escala en que se encuentre medida la variable; en este caso la escala está dada en años.

A la ordenada al origen, STATA la denota como `_cons` y aparece en el renglón siguiente al del coeficiente de regresión asociado con la única variable independiente. La interpretación directa (matemática) de este estimador es: el valor promedio que toma la variable de respuesta, en este caso `dmo`, cuando la variable independiente vale cero (en el ejemplo, cuando `edad=0`). En

el contexto de este problema, por supuesto, esto no tiene una interpretación plausible, pero en otro problema, el valor de la variable independiente igual a cero podría ser perfectamente posible. Más adelante, se verá que la estimación de $\sigma^2$ puede deducirse a partir del cuadro de análisis de la varianza (parte A); pero no se requiere para especificar la recta de regresión. De ahí que la ecuación del modelo estimado esté dada por:

$$\hat{\mu}_{\text{dmo}} = 0.6711 - 0.0052 \times \text{edad}$$

Hasta este momento no ha sido necesario hacer ningún supuesto distribucional; sin embargo, nótese que este método tiene ciertas limitaciones. Por ejemplo, si se obtiene una pendiente estimada de -0.0052, sería interesante generar un intervalo de confianza[1] para este parámetro, así como saber si, para fines prácticos, la pendiente de la recta estimada puede considerarse estadísticamente diferente de cero, es decir, es necesario determinar si hay evidencia de que la asociación detectada es real y no espuria (que no sea producto de la variabilidad intrínseca al muestreo). En conclusión, sería conveniente hacer inferencias estadísticas acerca de los parámetros. En la siguiente sección, se propone utilizar un método de estimación de los parámetros que incorpora un modelo probabilístico para el término de error. La ganancia será poder contestar estas preguntas; el costo, que el modelo sólo será aplicable a conjuntos de datos que cumplan, además, con este supuesto.



## ESTIMACIÓN DE LOS PARÁMETROS DE REGRESIÓN LINEAL SIMPLE POR MÁXIMA VEROSIMILITUD

Existen diversos procedimientos estadísticos para estimar parámetros, además del método de mínimos cuadrados. Un método utilizado frecuentemente es el de máxima verosimilitud,[2] ya que los estimadores obtenidos por este método tienen muchas propiedades estadísticas importantes cuando el tamaño de muestra es grande. Sin embargo, en el caso de la regresión lineal donde se tiene el supuesto de normalidad, las propiedades de estos estimadores se cumplen aún con muestras pequeñas.* Los valores estimados obtenidos mediante este método, hacen que la probabilidad de haber observado esta muestra particular sea máxima; es decir, el método elige como estimadores (de entre todos los posibles valores que pueden tomar los parámetros), aquellos que hacen más creíbles los datos observados.

Como se mencionó con anterioridad, de acuerdo con el modelo (1), hay tres parámetros de interés: la ordenada al origen $\beta_0$, la pendiente $\beta_1$ y la varianza del error, $\sigma^2$. Los estimadores máximo verosímiles de $\beta_0$ y $\beta_1$ coinciden exactamente con los que se obtienen por el método de mínimos cuadrados. Los estimadores de los coeficientes de regresión tienen una distribución normal, por lo que el cálculo de intervalos de confianza y pruebas de hipótesis puede realizarse con relativa facilidad. En lo que se refiere a la varianza del error, el estimador† con mejores propiedades estadísticas está dado por:

* Esto se debe a que las distribuciones de los coeficientes de regresión son exactas y no asintóticas (distribuciones que se alcanzan sólo cuando el tamaño de muestra es grande).
† Este estimador es una modificación del estimador máximo verosímil para hacerlo insesgado.

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \frac{SCE}{n-2}$$

Es interesante notar que este estimador es la suma de cuadrados que se minimiza con el procedimiento de mínimos cuadrados, dividida entre *n*-2; por lo tanto, el proceso de estimación busca que el error sea lo más pequeño posible. Se denotará como *SCE* a la suma de cuadrados del error.

## EL CUADRO DE ANÁLISIS DE LA VARIANZA

Después de obtener la ecuación de la recta de regresión, es necesario evaluarla para determinar si describe adecuadamente la relación entre las dos variables y si puede utilizarse con fines de predicción y/o estimación. Para ello, es muy útil analizar cómo se descompone la variabilidad de *Y*.

El modelo más simple que resume la información de *Y* es su media (modelo: $Y = \overline{y}$). No obstante, este modelo tan simple no incorpora la variable *X*; por ello, una forma de evaluar la calidad del modelo que contiene a *X* consiste en compararlo contra la versión más sencilla posible: $Y = \overline{y}$. Si el modelo que incorpora la variable *X* es estadísticamente "mejor" que el que la ignora, será evidencia de que la información que aporta *X* a la comprensión de *Y* es importante y, por ello, deberá incluirse en el modelo. La siguiente identidad muestra que la desviación total (diferencia entre el valor observado $y_i$ y la media $\overline{y}$) puede descomponerse en la suma de dos diferencias: una desviación que se explica por el modelo y otra que no (figura 6).



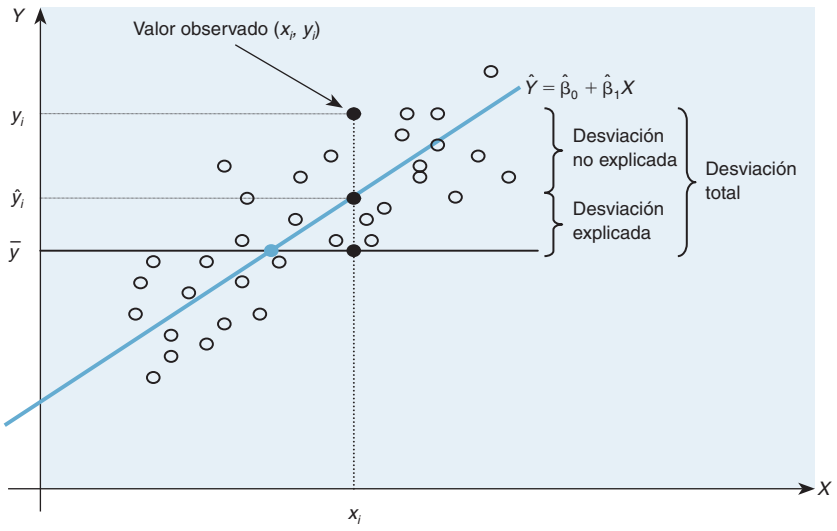



**Figura 6** **Descomposición de la variabilidad total de *Y***

$$\underbrace{y_i - \bar{y}}_{\text{desviación total}} = \underbrace{\left(\hat{y}_i - \bar{y}\right)}_{\text{desviación explicada por el modelo}} + \underbrace{\left(y_i - \hat{y}_i\right)}_{\text{desviación no explicada por el modelo}} \tag{3}$$

La desviación explicada por el modelo (diferencia entre el valor predicho por el modelo de regresión lineal y $\bar{y}$) es la ganancia en el conocimiento de $Y$, gracias al modelo de regresión lineal que incluye la variable explicativa $X$. La desviación no explicada por el modelo es el residuo. Si se elevan al cuadrado ambos lados de la identidad (3) y se simplifica algebraicamente, se obtiene la siguiente identidad:

$$\underbrace{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}_{\text{variación total}} = \underbrace{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}_{\text{variación explicada por la regresión}} + \underbrace{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}_{\text{variación de los residuos (no explicada)}} \tag{4}$$



Las siglas *SCT* y *SCR* denotarán la suma de cuadrados total y la suma de cuadrados debido a la regresión, respectivamente. Al reescribir la identidad (4) en términos de sumas de cuadrados, se tiene que:

$$SCT = SCR + SCE$$

La importancia de la identidad (4) radica en que la variabilidad de $Y$ puede descomponerse en dos partes: la variabilidad explicada por la regresión y la variabilidad de los residuos; es decir, una parte de la variabilidad de $Y$ se explica al controlar por la variable $X$, mientras que hay una parte de la variabilidad de $Y$ que el modelo propuesto no es capaz de explicar; sin embargo, recuérdese que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ se construyeron de tal forma que esta variabilidad no explicada fuera mínima. Sin embargo, aunque la *SCE* sea mínima, no tiene por qué ser pequeña. Intuitivamente puede decirse que, si la *SCE* es muy chica en relación con la variabilidad total, entonces la variable predictora $X$ explica mucho de la variabilidad de $Y$, y en este sentido, se considera el modelo de regresión como un buen modelo. Por otro lado, si la *SCE* fuera cero, significaría que todas las observaciones caen exactamente en la recta de regresión y que el ajuste es perfecto, lo cual en la práctica es casi imposible que suceda.

Los resultados de un análisis de regresión se resumen, de forma general, en lo que se conoce como *cuadro de análisis de la varianza* (cuadro I). Se le da este nombre porque muestra cómo se descompone la variabilidad de $Y$.

El cuadro de análisis de la varianza está dividido en renglones y columnas. Los renglones muestran las fuentes de la variabilidad total de $Y$: la variabilidad explicada por el modelo de regresión y la variabilidad residual. En las columnas se muestran las sumas de cuadrados, los gra-

**Cuadro I Análisis de la varianza para regresión lineal simple**

| *Fuente de variación* | *Sumas de cuadrados* | *Grados de libertad* | *Cuadrados medios (CM)* | $F_{obs.}$ | *Valor* p |
|---|---|---|---|---|---|
| Modelo (regresión) | $SCR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ | 1 | $CMR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ | $\frac{CMR}{CME}$ | $P[F_{1,n-2} > F_{obs}]$ |
| Error (residuos) | $SCE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ | *n*-2 | $CME = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}$ | | |
| Total | $SCT = \sum_{i=1}^{n} (y_i - \bar{y})^2$ | *n*-1 | | | |

dos de libertad y los cuadrados medios (cocientes de las sumas de cuadrados entre los grados de libertad). La prueba de hipótesis reportada en el cuadro de análisis de la varianza es la prueba de hipótesis que evalúa si la pendiente es igual o diferente a cero.

$$H_0: \beta_1 = 0 \ \ vs. \ \ H_1: \beta_1 \neq 0$$

donde $H_0$ es la hipótesis nula y $H_1$ es la alterna.

El estadístico de prueba utilizado sigue una distribución de probabilidad $F$ con 1 y $n$-2 grados de libertad. El resultado de esta prueba es de gran utilidad. Rechazar la hipótesis nula es evidencia estadística de que $X$ proporciona información significativa para predecir $Y$ (figura 7a) o bien que tal vez podría ajustarse un mejor modelo, pero que definitivamente tiene un componente lineal (figura 7b). Si, por el contrario, no se rechaza la hipótesis nula, entonces hay evidencia de que la variable independiente $X$ no ayuda a predecir $Y$ (figura 7c) o que la verdadera asociación entre $X$ y $Y$ no es lineal (figura 7d).

A diferencia de otros paquetes estadísticos, en STATA, la prueba global no viene reportada directamente en el cuadro tradicional de análisis de la varianza (parte A), sino en la parte B. En este caso, el valor de la estadística de prueba es $F(1, 432) = 147.36$ y el valor $p$ asociado a la prueba es muy pequeño ($p = 0.0000$), por lo que se llega a la conclusión de que la evidencia muestral apoya que la `edad` proporciona información significativa para predecir la densidad mineral ósea (`dmo`).

En el cuadro de análisis de la varianza (parte A), puede obtenerse el estimador de la varianza del error $\hat{\sigma}^2$, el cual es el valor del cuadrado medio del error ($CME$=0.004). El estimador de la varianza de `dmo` también puede encontrarse en este cuadro, en el renglón correspondiente a la fuente de variación total y en la columna de los cuadrados medios ($S^2_{dmo} = 0.0056$).

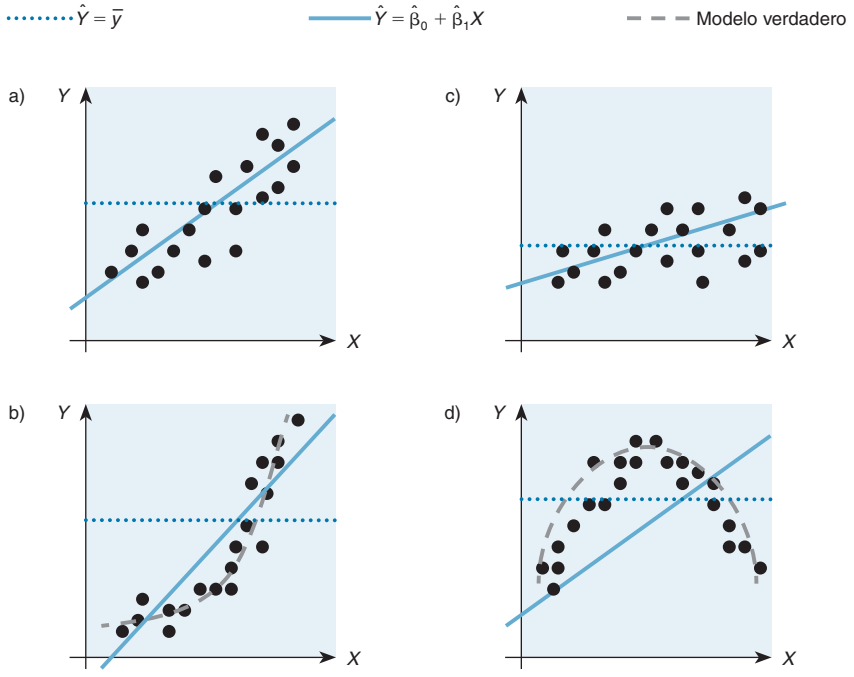



**Figura 7** **Interpretación de la prueba de hipótesis para $\beta_1$**

## COEFICIENTE DE DETERMINACIÓN

Una estadística de interés en el análisis de regresión lineal es el coeficiente de determinación que está dado por la siguiente ecuación:

$$R^2 = \frac{SCT - SCE}{SCT} = \frac{SCR}{SCT}$$

El coeficiente de determinación estima qué proporción de la variabilidad de *Y* se explica por medio de la regresión lineal de *Y* sobre *X*. Como es una proporción, entonces toma valores entre cero y uno, y en la práctica se expresa como porcentaje. Así, un valor cercano a uno indica que una gran proporción de la variabilidad de *Y* se explica al controlar por la variable *X*; mientras que un valor cercano a cero indica que la variable *X* ayuda poco a explicar la variabilidad de *Y*. El coeficiente de determinación coincide con el cuadrado del coeficiente de correlación de Pearson entre *X* y *Y*. Es importante mencionar que el coeficiente de determinación no es una medida que indique si el modelo lineal es apropiado o no. El supuesto de linealidad se verificará por medio de los residuos del modelo (véase el capítulo de regresión lineal múltiple).

Retomando el ejemplo de dmo se observa, en la parte B de la salida de STATA, que el coeficiente de determinación $R^2$ es 0.2543. Esto significa que, aproximadamente, la edad explica el 25% de la variabilidad total de la densidad mineral ósea de las mujeres participantes en el estudio (dmo).

## INFERENCIAS SOBRE LA ORDENADA AL ORIGEN Y LA PENDIENTE

Para poder hacer inferencias sobre los parámetros $\beta_0$ y $\beta_1$, es necesario el supuesto de normalidad de los errores. Bajo este supuesto, los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ tienen una distribución normal con medias $\beta_0$ y $\beta_1$ y su varianza depende del parámetro desconocido $\sigma^2$.

### Intervalos de confianza

Para calcular los intervalos de confianza de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$, es necesario calcular sus errores estándar. Debido a que las varianzas de $\hat{\beta}_0$ y $\hat{\beta}_1$ dependen del parámetro desconocido $\sigma^2$, es necesario estimar dicho parámetro. Así como el intervalo de confianza de la media de una variable normal, cuya varianza es desconocida, involucra la distribución muestral *t*, los intervalos de confianza de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ involucran esta misma distribución. Los intervalos de confianza para los parámetros $\beta_0$ y $\beta_1$ están dados por las siguientes ecuaciones:

$$\hat{\beta}_0 \pm t_{n-2}^{1-\alpha/2} \times \text{error estándar de } \hat{\beta}_0$$
$$\hat{\beta}_1 \pm t_{n-2}^{1-\alpha/2} \times \text{error estándar de } \hat{\beta}_1$$

donde $t_{n-2}^{1-\alpha/2}$ es el cuantil 1- $\alpha/2$ de una distribución *t* con *n*-2 grados de libertad.

Los libros especializados en regresión lineal[2-4] presentan las fórmulas para el cálculo de los errores estándar. En la parte C de la salida de STATA puede verse que, para cada coeficiente de regresión, se presenta el error estándar del estimador así como su intervalo de confianza de 95 por ciento.

### Prueba de hipótesis para el coeficiente de regresión $\beta_1$

La prueba de hipótesis más importante de un modelo de regresión lineal simple consiste en determinar si el coeficiente de regresión asociado con la única variable independiente es significativamente diferente de cero. Puesto que el modelo simple incluye sólo una variable independiente, esta prueba de hipótesis es equivalente a la que se reporta en el cuadro de análisis de la varianza; las conclusiones derivadas de esta prueba ya han sido previamente explicadas con ayuda de la figura 7.

En los diferentes paquetes estadísticos, el valor *p* asociado con la prueba de hipótesis sobre la significancia del coeficiente se reporta al lado del coeficiente mismo. Como era de esperarse, el valor *p* asociado a la prueba de hipótesis del coeficiente de regresión $\beta_1$ es el mismo que el del cuadro de análisis de la varianza, ya que ambas pruebas se refieren al único coeficiente de regresión en juego en un modelo de regresión lineal simple. Sin embargo, hay que notar que la esta-

dística de prueba del cuadro de análisis de la varianza tiene una distribución F y no una *t*. Esto se debe a que puede demostrarse matemáticamente que para $\nu$ grados de libertad,

$$F_{1,\nu}^{1-\alpha} = \left(t_{\nu}^{1-\alpha/2}\right)^2$$

Retomando el ejemplo, el estimador del coeficiente de regresión asociado con `edad` ($\hat{\beta} = -0.0052$) es significativamente diferente de cero ($p = 0.000$); esto indica que la disminución promedio en `dmo` es estadísticamente diferente de cero. Restaría determinar si esta disminución es biológicamente relevante.

**Prueba de hipótesis para la ordenada al origen $\beta_0$**

A diferencia de la prueba de hipótesis de la pendiente, la prueba de hipótesis para la ordenada al origen es de muy poco interés en la mayoría de las aplicaciones. La razón es que, por lo general, las observaciones no están muy cerca del origen e, incluso, en ocasiones la variable predictora *X* no toma el valor de cero (por ejemplo, la edad de mujeres posmenopáusicas no puede ser cero). Matemáticamente se requiere estimar la ordenada al origen para poder determinar la ecuación de regresión; pero es importante recordar que no necesariamente debe tener interpretación biológica plausible.

## PRUEBA t PARA DOS MUESTRAS INDEPENDIENTES COMO CASO PARTICULAR DE LA REGRESIÓN LINEAL



Hasta este momento, se ha asumido que la variable independiente X es continua; pero no tiene por qué serlo. En particular, cuando la variable independiente es dicotómica, puede demostrarse que el modelo de regresión lineal simple se reduce a una prueba *t* para dos muestras independientes con varianzas iguales. Si la variable dicotómica se codifica como cero y uno, entonces los parámetros $\beta_0$ y $\beta_1$ tienen interpretaciones muy útiles. Por ejemplo, la ordenada al origen es el promedio de *Y*, cuando la variable independiente es igual a cero; es decir para el grupo que está codificado como cero. Este grupo suele llamarse de referencia. Si el modelo de regresión incluye la ordenada al origen, como sucede en la mayoría de las aplicaciones, entonces el coeficiente de regresión asociado a *X* es la estimación de la diferencia de los promedios entre los dos grupos. Esta diferencia entre los promedios también se interpreta como la diferencia promedio. Como consecuencia, la prueba de hipótesis asociada con este parámetro tiene una interpretación muy útil. Si esta diferencia es significativamente diferente de cero, entonces podemos concluir que hay evidencia estadística para decir que las medias de los dos grupos son diferentes. Un punto importante a resaltar es que los valores *p* reportados en el análisis de regresión corresponden a pruebas de dos colas. Entonces, si la hipótesis del investigador es de una sola cola, hay que dividir entre dos el valor *p*, mostrado en la salida del modelo de regresión lineal.

Retomando el ejemplo, una pregunta de interés para los investigadores es si la `dmo` es mayor en las mujeres que usan terapia hormonal de reemplazo, comparada con las que no la usan, o sea que la prueba de hipótesis de interés es:

$$H_0: \mu_{\texttt{thr}} = \mu_{\texttt{nthr}} \quad vs. \quad H_1: \mu_{\texttt{thr}} > \mu_{\texttt{nthr}}$$

donde thr indexa al grupo que usa terapia hormonal de reemplazo, y nthr al que no la usa. Para comparar las medias de dmo entre ambas poblaciones, puede usarse una prueba *t* para dos muestras independientes;[1] pero primero habría que determinar si la distribución de dmo en cada grupo es normal y si las varianzas son iguales en ambos grupos. En STATA, los comandos **ttest**, **swilk** y **sdtest** hacen estas pruebas respectivamente.

Suponiendo que se llega a la conclusión de que las varianzas son iguales y las poblaciones son normales, se procede a realizar una prueba *t* para comparar las medias de dos muestras independientes con varianzas iguales, o bien, puede utilizarse un modelo de regresión lineal, ya que esta prueba puede verse como un caso particular de este modelo.

En este ejemplo, la variable dependiente es dmo y la variable independiente es el uso o no de terapia hormonal de reemplazo (thr) codificada como uno y cero, respectivamente; es decir, el grupo de referencia es el grupo de mujeres que no usan thr. Se utiliza el comando **tab** para estimar la prevalencia de uso de thr en la muestra de estudio (22.6%):

```
tab  thr
     Uso de |
    terapia |
   hormonal |
  1=sí 0=no |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        336       77.42       77.42
          1 |         98       22.58      100.00
------------+-----------------------------------
      Total |        434      100.00
```

Con el objetivo de estimar el promedio de dmo en cada categoría de thr, se utiliza la opción **summ(**dmo**)** del comando **tab**, ya que éste proporciona estadísticas descriptivas de la variable dmo para cada categoría de la variable que se esté tabulando (en este caso, thr).

```
tab  thr, summ(dmo)

     Uso de |
    terapia |
   hormonal |  Summary of Densidad mineral osea
  1=si 0=no |        Mean   Std. Dev.       Freq.
------------+------------------------------------
          0 |   .38595536   .07671608         336
          1 |   .39676531   .06908023          98
------------+------------------------------------
      Total |   .38839631   .07511887         434
```

A continuación, se verifica si la variable dmo tiene una distribución normal en cada uno de los grupos definidos por la variable thr, por medio de la prueba de bondad de ajuste Shapiro-Wilk.[5] La hipótesis nula de esta prueba es que la distribución probabilística de la muestra en es-

tudio no difiere significativamente de una normal, por lo que valores *p* "grandes" apoyarán la normalidad de la muestra, mientras que valores *p* pequeños sugerirán diferencias significativas entre la distribución probabilística de la muestra y una distribución normal:

```
by thr: swilk dmo
```

```
-------------------------------------------------------------------------------
-> thr = 0

                    Shapiro-Wilk W test for normal data
    Variable |    Obs        W           V          z      Prob>z
-------------+-------------------------------------------------
         dmo |    336    0.99677      0.762     -0.642   0.73956



-------------------------------------------------------------------------------
-> thr = 1

                    Shapiro-Wilk W test for normal data
    Variable |    Obs        W           V          z      Prob>z
-------------+-------------------------------------------------
         dmo |     98    0.98597      1.139      0.288   0.38662
```



Los resultados de estas pruebas de hipótesis sugieren que el supuesto de normalidad se cumple satisfactoriamente. Ahora, compárense las varianzas de la variable `dmo` en cada uno de los grupos definidos por `thr`:

```
sdtest  dmo, by(thr)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     336    .3859554    .0041852    .0767161    .3777228     .394188
       1 |      98    .3967653    .0069782    .0690802    .3829156     .410615
---------+--------------------------------------------------------------------
combined |     434    .3883963    .0036058    .0751189    .3813092    .3954834
------------------------------------------------------------------------------

                         Ho: sd(0) = sd(1)

             F(335,97) observed   = F_obs             =     1.233
             F(335,97) lower tail = F_L   = 1/F_obs =     0.811
             F(335,97) upper tail = F_U   = F_obs   =     1.233
```

```
   Ha: sd(0) < sd(1)          Ha: sd(0) ~= sd(1)          Ha: sd(0) > sd(1)
P < F_obs = 0.8905     P < F_L + P > F_U = 0.1999     P > F_obs = 0.1095
```

Este valor *p* de 0.2 sugiere que no existe evidencia muestral que apoye la diferencia entre las varianzas, por lo que puede considerarse que el supuesto de homogeneidad de varianzas también se cumple.

Después de verificar estos supuestos, sólo falta analizar la independencia entre las observaciones para poder proceder a realizar una prueba *t* para comparar las medias de `dmo` entre estas dos muestras. Recuérdese que se trata de un estudio transversal en mujeres trabajadoras del IMSS en Morelos. De acuerdo a la forma de seleccionar la muestra, no existen mediciones repetidas, ni agregación familiar ni ninguna estructura que sugiera correlación entre las participantes, por lo que se verificaron todos los supuestos para poder proseguir con la prueba *t*.

La siguiente salida muestra la prueba *t* para dos muestras independientes con varianzas iguales. Puede observarse que la diferencia promedio en `dmo` es 0.01 g/cm$^2$ mayor en las mujeres que usan terapia hormonal de reemplazo, comparada con las que no la usan, pero que esta diferencia no es estadísticamente significativa (valor $p = 0.1052$):

```
ttest dmo, by( thr)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     336    .3859554    .0041852    .0767161    .3777228     .394188
       1 |      98    .3967653    .0069782    .0690802    .3829156     .410615
---------+--------------------------------------------------------------------
combined |     434    .3883963    .0036058    .0751189    .3813092    .3954834
---------+--------------------------------------------------------------------
    diff |           -.0108099    .0086183               -.0277491    .0061292
------------------------------------------------------------------------------
Degrees of freedom: 432

                    Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =  -1.2543                 t =  -1.2543                 t =  -1.2543
   P < t =   0.1052           P > |t| =   0.2104           P > t =   0.8948
```

Ahora, compárense los resultados con los obtenidos a partir del modelo de regresión lineal. Puede verse que el estimador de la ordenada al origen (0.3859554) es el promedio de densidad mineral ósea de las mujeres sin `thr` (grupo de referencia) y el estimador del coeficiente de regresión asociado a la variable `thr` (0.0108099) es la diferencia de densidad mineral ósea entre los promedios de los dos grupos. Esta diferencia es positiva e indica que el promedio de `dmo` de

las mujeres que usan `thr` es mayor. De hecho, la densidad mineral ósea promedio en las mujeres que usan `thr` es la suma de la ordenada al origen más el coeficiente de regresión; es decir 0.3859554 + 0.0108099 = 0.3967653 g/cm$^2$:

```
regress dmo thr

      Source |       SS       df       MS              Number of obs =     434
-------------+------------------------------           F(  1,   432) =    1.57
       Model |  .008865901     1  .008865901           Prob > F      =  0.2104
    Residual |  2.43448593   432  .005635384           R-squared     =  0.0036
-------------+------------------------------           Adj R-squared =  0.0013
       Total |  2.44335184   433  .005642845           Root MSE      =  .07507

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .0108099   .0086183     1.25   0.210    -.0061292    .0277491
       _cons |   .3859554   .0040954    94.24   0.000      .377906    .3940047
------------------------------------------------------------------------------
```

Como se mencionó previamente, los valores *p* reportados en el análisis de regresión corresponden a pruebas de dos colas. Sin embargo, la hipótesis del investigador era de una sola cola, por lo que hay que dividir entre dos el valor *p* correspondiente a la prueba de hipótesis asociada con el coeficiente de regresión de `thr` (0.2104), para obtener el valor *p* de una cola, que es exactamente el reportado del lado izquierdo del comando **`ttest`**.



## VARIABLES INDEPENDIENTES POLITÓMICAS

Cuando la variable independiente es politómica (variable con más de dos categorías), deben construirse variables dicotómicas (*dummies* o indicadoras) asociadas con cada una de las categorías de la variable para introducirlas al modelo, en lugar de la variable original. Esto se debe a que los números que representan las categorías son etiquetas arbitrarias que no necesariamente representan una escala. Aun en las variables ordinales, en donde existe un orden en las categorías, el número asignado a cada una de ellas refleja el orden mismo de la escala, mas no es interpretable su magnitud. Si la variable politómica $X$ tiene $k$ categorías, entonces se pueden construir $k$ variables dicotómicas ($X_1, X_2, \ldots, X_k$, donde el subíndice indexa el número de categoría). Es muy importante notar que los valores asignados para identificar cada una de las categorías en las variables indicadoras determinan el análisis y la interpretación.

La forma más utilizada para codificar las variables indicadoras es con ceros y unos (la variable toma el valor de uno si la observación pertenece a la categoría correspondiente, y cero si no pertenece). En los modelos de regresión que incluyen la ordenada al origen, es suficiente introducir $k$-1 variables indicadoras, porque incluirlas todas es redundante. Con $k$-1 variables pode-

mos describir completamente la pertenencia a cualquiera de las *k* categorías, ya que los sujetos en los que todas estas variables sean iguales a cero, necesariamente, serán sujetos que pertenezcan a la categoría de referencia. Al codificar de esta forma, el coeficiente de regresión asociado con cada categoría se interpreta como la diferencia, entre el promedio de *Y* en la categoría correspondiente y el promedio de *Y* en la categoría de referencia. La ordenada al origen se interpreta, entonces, como el promedio de *Y* para el grupo de referencia.

## ANÁLISIS DE VARIANZA DE UNA SOLA VÍA, COMO CASO PARTICULAR DEL MODELO DE REGRESIÓN LINEAL

Otro caso particular del modelo de regresión lineal es el análisis de varianza de una sola vía, conocida como ANOVA. Se dice que es de una sola vía porque sólo hay una variable que clasifica los grupos. Aunque la variable que determina los *k* grupos es una sola, en el modelo de regresión lineal se expresa con *k*-1 variables dicotómicas. La prueba global del análisis de varianza de una sola vía es equivalente a la prueba *F*, reportada en el cuadro de análisis de la varianza de una regresión lineal. La estadística de prueba para evaluar si las medias de los *k* grupos son iguales tiene *k*-1 grados de libertad, ya que se requieren *k*-1 igualdades para probar la hipótesis nula ($H_0$: $\mu_1=\mu_2=\ldots=\mu_{k-1}=\mu_k$, donde $\mu_i$ es la media del grupo i).

Las pruebas de hipótesis reportadas junto con los coeficientes de regresión evalúan si la diferencia entre la media de *Y*, del grupo al que corresponde ese coeficiente, y la del grupo de referencia es significativamente diferente a cero. En un modelo de regresión con ordenada al origen y *k*-1 variables dicotómicas codificadas con cero y uno, los coeficientes de regresión son diferencias promedio de *Y*, respecto al grupo de referencia. Esto implica que, si un coeficiente es significativamente diferente de cero, entonces la media de *Y* del grupo al que corresponde ese coeficiente no es igual a la media del grupo de referencia. Es importante notar que por medio de un modelo de regresión lineal, sólo se obtienen las pruebas de las diferencias entre cada grupo y el de referencia; pero no entre todos los grupos (como es lo usual en la mayoría de los paquetes estadísticos que hacen ANOVA). Para obtener estas pruebas habría que pedirlas explícitamente.

Volviendo al ejemplo de `dmo`, si al investigador sólo le interesa comparar el promedio de `dmo` entre tres grupos correspondientes al estado de nutrición de las mujeres (de acuerdo con tres categorías del índice de masa corporal), entonces podría hacerse un análisis de varianza ANOVA de una sola vía con el comando **`oneway`** de STATA. Análogamente, se puede hacer un modelo de regresión lineal múltiple incorporando dos variables indicadoras, `imc2` e `imc3`, para sobrepeso (`imc` $\geq 25$ e `imc` $< 30$) y obesidad (`imc` $\geq 30$), respectivamente.

Para crear las variables indicadoras, se utiliza la opción **`gen`** del comando **`tab`** que, automáticamente, permite generar las correspondientes a las categorías de la variable que se esté tabulando. Se utiliza el comando **`summ`** para mostrar que las variables dicotómicas generadas toman los valores 0 y 1:

```
tab imc_3cat, gen(imc)

        IMC |
categorizado|
    en 3    |
 (Puntos de |
 Corte 25 y |
        30) |      Freq.     Percent        Cum.
------------+-----------------------------------
   Adecuado |        104       23.96       23.96
  Sobrepeso |        189       43.55       67.51
   Obesidad |        141       32.49      100.00
------------+-----------------------------------
      Total |        434      100.00


. summ  imc1 imc2 imc3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        imc1 |       434    .2396313    .4273511          0          1
        imc2 |       434    .4354839    .4963924          0          1
        imc3 |       434    .3248848    .4688723          0          1
```



El siguiente comando permite conocer el promedio de `dmo` en cada uno de los grupos definidos por la variable `imc_3cat`:

```
tab  imc_3cat, summ(dmo)

        IMC |
categorizado|
     en 3   |
 (Puntos de |
 Corte 25 y |  Summary of Densidad mineral osea
        30) |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   Adecuado |   .35777885   .07176133         104
  Sobrepeso |   .39262963   .07322953         189
   Obesidad |   .40530496   .07385216         141
------------+------------------------------------
      Total |    .38839631   .07511887         434
```

A continuación, se muestran las dos salidas correspondientes al comando **oneway** y al **regress**; puede verificarse que la prueba global para determinar si la media de algún grupo es diferente es la misma. En este caso, la prueba global es significativa ($p = 0.000$) e indica que, por lo menos, la media de un grupo es diferente a las demás. A partir de las estadísticas descriptivas, puede comprobarse que el estimador de la ordenada al origen del modelo de regresión (0.3578) es la dmo promedio del grupo de referencia (grupo 1: peso adecuado). Los estimadores de los coeficientes de las variables indicadoras son las diferencias entre los promedios en dmo para los grupos 2 y 3 respecto al promedio del grupo de referencia. Si se suman las diferencias, imc2 e imc3, al estimador de la ordenada al origen, se obtienen los promedios de dmo para las mujeres con sobrepeso y obesidad, respectivamente. Por ejemplo, la dmo promedio en mujeres con sobrepeso es 0.3926296 g/cm$^2$ (que se obtiene al sumar 0.0348508 con 0.3577788). Los valores *p* asociados a los coeficientes de regresión permiten concluir que las medias de dmo para los grupos 2 y 3 son significativamente diferentes de la media de dmo del grupo 1:

```
oneway dmo imc_3cat
                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups       .141191962      2   .070595981     13.22     0.0000
 Within groups       2.30215987    431   .005341438
------------------------------------------------------------------------
    Total            2.44335184    433   .005642845

Bartlett's test for equal variances:  chi2(2) =   0.0999  Prob>chi2 = 0.951


regress dmo imc2 imc3

      Source |       SS       df       MS              Number of obs =     434
-------------+------------------------------           F(  2,   431) =   13.22
       Model |  .141191962     2  .070595981           Prob > F      =  0.0000
    Residual |  2.30215987   431  .005341438           R-squared     =  0.0578
-------------+------------------------------           Adj R-squared =  0.0534
       Total |  2.44335184   433  .005642845           Root MSE      =  .07309

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        imc2 |   .0348508   .0089231     3.91   0.000     .0173126     .052389
        imc3 |   .0475261   .0094468     5.03   0.000     .0289585    .0660937
       _cons |   .3577788   .0071666    49.92   0.000      .343693    .3718647
------------------------------------------------------------------------------
```

Para probar si la media del grupo 2 es diferente a la del grupo 3, habría que pedir explícitamente la prueba de hipótesis, ya sea con el comando **test** o con el comando **lincom**. Con ambos comandos se especifica la hipótesis nula, sólo que con el comando **test**, explícitamente se pide que un coeficiente sea igual que el otro ($H_0$: $\mu_2 = \mu_3$), mientras que con el comando **lincom**, se especifica una combinación lineal que quiere probar que es igual a cero ($H_0$: $\mu_2 - \mu_3 = 0$). Estos dos comandos hacen exactamente la misma prueba, sólo que uno utiliza una estadística de prueba *t* y el otro una *F*; sin embargo, recuérdese que la estadística *t* elevada al cuadrado tiene una distribución *F*. Además, el comando **lincom** genera un intervalo de confianza para el estimador de la combinación lineal especificada. Otra forma de hacerlo es mediante un nuevo modelo de regresión, en el que se utilice como referencia al grupo 2 o 3. A continuación se muestran las salidas donde puede verificarse la equivalencia de estas pruebas de hipótesis:

```
test imc2 = imc3

 ( 1)  imc2 - imc3 = 0.0

       F(  1,   431) =    2.43
            Prob > F =    0.1198

lincom imc2 - imc3

 ( 1)  imc2 - imc3 = 0.0

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0126753   .0081329    -1.56   0.120    -.0286604    .0033098
------------------------------------------------------------------------------

regress dmo imc1 imc2

      Source |       SS       df       MS              Number of obs =     434
-------------+------------------------------           F(  2,   431) =   13.22
       Model |  .141191962     2  .070595981           Prob > F      =  0.0000
    Residual |  2.30215987   431  .005341438           R-squared     =  0.0578
-------------+------------------------------           Adj R-squared =  0.0534
       Total |  2.44335184   433  .005642845           Root MSE      =  .07309

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        imc1 |  -.0475261   .0094468    -5.03   0.000    -.0660937   -.0289585
        imc2 |  -.0126753   .0081329    -1.56   0.120    -.0286604    .0033098
       _cons |    .405305   .0061549    65.85   0.000     .3932077    .4174023
------------------------------------------------------------------------------
```

## REFERENCIAS


1. Pagano M, Gauvreau K. Fundamentos de bioestadística. México: Thomson Learning, 2001.
2. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Applied regression analysis and other multivariable methods. Nueva York: Duxbury Press, 1998.
3. Draper NR, Smith H. Applied regression analysis. Wiley Series in Probability and Statistics. Washington: A Wiley-Interscience Publication, 1998.
4. Montgomery DC, Peck EA. Introducción al análisis de regresión lineal. México: CECSA, 2002.
5. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965; 52:591-611.

# XIII Regresión lineal múltiple

*Daniela Sotres Álvarez*
*Martha María Téllez Rojo Solís*

## INTRODUCCIÓN

Como se mencionó en el capítulo anterior, el análisis de regresión lineal múltiple es un método estadístico que evalúa la asociación entre una variable de respuesta continua $Y$ con dos o más variables explicativas $X_1, X_2,\ldots, X_k$ ($k \geq 2$). Las variables explicativas pueden ser continuas o politómicas, o una mezcla de ambas. Un modelo de regresión lineal simple no es más que un caso particular del modelo múltiple cuando $k = 1$.

Es importante señalar la distinción estadística entre los términos múltiple y multivariado. En la práctica, e incluso en la literatura sobre el tema, llegan a emplearse indistintamente, aunque en realidad describen modelos diferentes. En un modelo de regresión lineal múltiple se estudia la asociación entre las variables explicativas $X_1, X_2,\ldots, X_k$ y una única variable dependiente $Y$. Por otro lado, en un modelo de regresión lineal multivariado, se estudia la asociación entre este grupo de variables independientes y dos o más variables de respuesta $Y_1, Y_2,\ldots, Y_q$ observadas en un mismo individuo. Este último es un modelo estadístico más avanzado que no se cubre en textos básicos de bioestadística.[1,2]



## EL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

Para estimar el modelo de regresión lineal múltiple se requiere de una muestra de $n$ sujetos, con información de $k$ variables independientes $X_1, X_2, \ldots, X_k$ y de una variable dependiente Y, considerada "consecuencia" o respuesta del conjunto de variables independientes. Indexaremos con $i$ al sujeto y con $j$ a la variable independiente. Así, $x_{ij}$ representa el valor observado de la variable independiente $j$ para el sujeto $i$. El modelo de regresión lineal sólo utiliza sujetos que no tienen datos faltantes en ninguna de las variables involucradas en el modelo.

El modelo teórico de regresión lineal múltiple con $k$ variables predictivas $X_1, X_2,\ldots, X_k$ para el $i$-ésimo sujeto se especifica como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, 2, \ldots, n$$

en donde $\varepsilon_i$ es el término de error que tiene una distribución normal con media cero y varianza $\sigma^2$. Los parámetros a estimar en un modelo de regresión lineal múltiple son la ordenada al origen, los $k$ coeficientes asociados a las $k$ variables predictivas y la varianza del error ($\sigma^2$). Aquí podemos ver que para el caso en que $k = 1$, este modelo es exactamente el modelo de regresión li-

neal simple que presentamos en el capítulo anterior. A los parámetros $\beta_0, \beta_1, \ldots, \beta_k$ se les conoce como coeficientes de regresión.

Igual que en el modelo de regresión lineal simple, los parámetros pueden estimarse tanto con el método de mínimos cuadrados como con el de máxima verosimilitud. En ambos métodos se requieren los supuestos descritos en el capítulo anterior: independencia de las observaciones, linealidad en la asociación entre la variable de respuesta y cada una de las variables independientes, homoscedasticidad y, para el caso específico del método de máxima verosimilitud, el supuesto adicional de normalidad del término de error. Los estimadores derivados por ambos métodos son iguales. Sin embargo, el de mínimos cuadrados, por un lado, al no imponer ningún supuesto distribucional sobre el término de error, limita su aplicación a la estimación puntual de los parámetros; por el otro, el método de máxima verosimilitud, al imponer el supuesto de normalidad en los errores, permite ampliar su aplicación para hacer inferencias estadísticas sobre los parámetros, como calcular intervalos de confianza y hacer pruebas de hipótesis. La deducción matemática y las propiedades estadísticas de estos estimadores pueden consultarse en libros especializados en el tema.[3] Cuando alguno de los supuestos no se cumple, es necesario recurrir a estrategias estadísticas alternativas.[4]

## INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESIÓN



Veamos un ejemplo. Desea determinarse si el uso de terapia hormonal (`thr`) se asocia con la densidad mineral ósea (`dmo`). Se sabe que el metabolismo óseo se ve influenciado por algunas variables, además de la terapia hormonal, que deberían tomarse en consideración como potenciales confusores; las principales son: edad en años (`edad`), índice de masa corporal (`imc`), ingesta diaria de calcio en mg/día (`calcio`) y actividad física semanal, cuya unidad de medición está dada en equivalentes metabólicos denominados MET* (`actfis`).

Como es habitual, la información sobre el índice de masa corporal está dada por una variable continua, que resulta de dividir el peso (kg) entre el cuadrado de la talla (m) de cada participante. La riqueza de la información de una variable continua es algo que debe de aprovecharse; sin embargo, muy frecuentemente sucede que existen puntos de corte preestablecidos para ciertas variables, cuyo uso es ampliamente difundido entre la comunidad científica del tema particular, lo que hace indispensable utilizarlos. Desde el punto de vista estadístico, sería preferible iniciar con la variable continua, y que los mismos datos "hablaran" y sugirieran puntos de corte para la población de estudio particular. Sin embargo, esto impediría la comparabilidad con otros estudios, razón por la que frecuentemente prevalece el criterio epidemiológico y se adopta la categorización propuesta. Esto sucede con el índice de masa corporal, que se analizará de manera categórica por medio de la utilización de puntos de corte que atienden criterios preestablecidos: se considerará que una mujer tiene peso adecuado cuando su `imc` $< 25$; con sobrepeso, cuando esté entre 25 y 30 (`imc` $\geq 25$, `imc` $< 30$), y con obesidad cuando `imc` $\geq 30$.

* Un equivalente metabólico (MET) representa un múltiplo de la cantidad de oxígeno consumida en estado de reposo. Si al hacer cierto ejercicio una persona tiene un gasto energético de 10 MET, por ejemplo, significa que ha consumido 10 veces la cantidad de oxígeno que normalmente consumiría si estuviese en reposo.

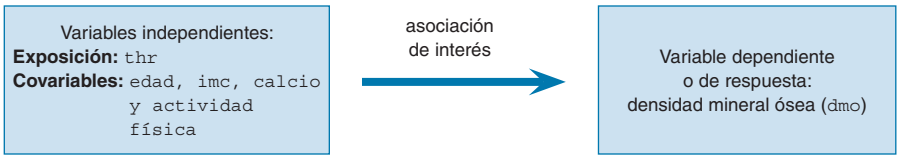



**Figura 1** **Asociación entre terapia hormonal de reemplazo (thr) y densidad mineral ósea, ajustada por covariables**

La asociación que se estudia y la clasificación de las variables involucradas se resumen en la figura 1.

Obsérvense algunas estadísticas descriptivas de las variables involucradas en el modelo. Debido a que la variable `thr` está codificada como uno si se usa terapia hormonal sustitutiva y como cero si no se usa, entonces la media aritmética equivale a la proporción de mujeres que usan `thr`:

```
summ  dmo thr edad calcio actfis imc

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         dmo |     434    .3883963    .0751189        .13        .591
         thr |     434    .2258065    .4185948          0          1
        edad |     434    54.16359    7.258567         41         80
      calcio |     430    1143.911    694.8332      38.05    5952.01
      actfis |     434    6.293362    10.06693          0         63
         imc |     434    28.43264    4.852104   17.06556   48.44444
```

Utilizando el comando **tab** con la opción **gen**, pueden generarse las variables indicadoras o *dummies* asociados a la variable categórica `imc_3cat`, que contiene las tres categorías antes descritas.

```
tab  imc_3cat, gen(imc)

        IMC |
categorizado|
       en 3 |
 (Puntos de |
 Corte 25 y |
        30) |      Freq.     Percent        Cum.
------------+-----------------------------------
   Adecuado |        104       23.96       23.96
  Sobrepeso |        189       43.55       67.51
   Obesidad |        141       32.49      100.00
------------+-----------------------------------
      Total |        434      100.00
```

En STATA, el comando para hacer una regresión lineal es **regress**, que va seguido de la variable dependiente y después de las variables independientes:

```
regress dmo thr edad imc2 imc3 calcio actfis
```

```
      Source |       SS       df       MS              Number of obs =     430
-------------+------------------------------           F(  6,   423) =   33.30
       Model |  .767103179     6   .12785053           Prob > F      =  0.0000
    Residual |  1.62405512   423  .003839374           R-squared     =  0.3208
-------------+------------------------------           Adj R-squared =  0.3112
       Total |   2.3911583   429  .005573796           Root MSE      =  .06196

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .0106175   .0072816     1.46   0.146    -.0036951      .02493
        edad |  -.0050006   .0004161   -12.02   0.000    -.0058185   -.0041828
        imc2 |   .0331696   .0076072     4.36   0.000     .0182169    .0481222
        imc3 |   .0447026   .0081155     5.51   0.000      .028751    .0606543
      calcio |   .0000104   4.34e-06     2.40   0.017     1.87e-06    .0000189
      actfis |   .0005848   .0003077     1.90   0.058    -.0000199    .0011896
       _cons |   .6118931   .0242392    25.24   0.000     .5642489    .6595374
------------------------------------------------------------------------------
```



A partir de esta salida, puede verse que la ecuación del modelo ajustado está dada por:

$$\hat{\texttt{dmo}} = 0.6118931 + 0.0106175 \times \texttt{thr} - 0.0050006 \times \texttt{edad} + 0.0331696 \times \texttt{imc2} + 0.0447026 \times \texttt{imc3} + 0.0000104 \times \texttt{calcio} + 0.0005848 \times \texttt{actfis}$$

se le pone *gorro* a la media de la variable dependiente dmo, para enfatizar que se trata del modelo estimado para la media de la variable observada dmo. Se usarán indistintamente los términos modelo estimado y modelo ajustado para referirse al modelo en donde ya se estimaron los coeficientes de regresión, utilizando la información muestral del estudio.

Los estimadores de los coeficientes de regresión se interpretan de la misma forma que en regresión lineal simple; pero hay que mencionar que "están ajustados" por las covariables incluidas en el modelo. En este ejemplo, el coeficiente de mayor interés es el asociado con la variable thr, porque es el que está indicando la magnitud de la diferencia ajustada entre la media de dmo en la población de mujeres que usa terapia hormonal y la media de dmo en la población que no. Esta diferencia es, en promedio, de 0.0106175 g/cm$^2$ en favor de la población de mujeres que usan thr, ajustando por las variables edad, índice de masa corporal politómica, ingesta de calcio y actividad física. Sabemos que la diferencia favorece a la población de mujeres que usan thr, porque el signo del coeficiente de regresión correspondiente es positivo, y la comparación se realiza entre las mujeres que sí usaron thr (thr = 1) *versus* las que no la usaron (thr = 0).

Ahora bien, cuando se dice "ajustando por las demás variables" significa que se compara el promedio de densidad mineral ósea entre la población de mujeres que usan terapia hormonal y la población de mujeres que no usan terapia hormonal, siempre y cuando tengan los mismos valores en las otras cuatro variables. En otras palabras, si se comparan dos poblaciones de mujeres de la misma edad, con un índice de masa corporal que las colocara en el mismo grupo de `imc_3cat`, que tuvieran una ingesta diaria de calcio idéntica y una actividad física que consumiera la misma cantidad de MET, entonces podría decirse que, en promedio, la mujer que utiliza `thr` tendrá una `dmo` superior a la que no la usa en 0.0106175 g/cm$^2$, como se mencionó anteriormente. Resumiendo, la idea de estimar el efecto de la variable independiente $X_j$ en las medias de la variable de respuesta, ajustando por las covariables incluidas en el modelo, consiste en comparar dos poblaciones que tienen exactamente los mismos valores en estas covariables; pero que difieren solamente en una unidad de la variable $X_j$.

Esta interpretación puede deducirse de las siguientes expresiones algebraicas en las que se observa que la diferencia estimada promedio en la variable de respuesta $Y$ entre dos poblaciones, una cuyo valor observado en la variable $X_1$ fue $(x_1 + 1)$, y otra cuyo valor observado fue $x_1$, está dada por el estimador del coeficiente de regresión asociado con la variable $X_1$, siempre y cuando las demás variables $X_2, X_3, \ldots, X_k$ tomen exactamente el mismo valor para ambos sujetos:

$$\hat{Y}_{x_1} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

$$\hat{Y}_{x_1+1} = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + 1) + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

$$\hat{Y}_{x_1+1_1} - \hat{Y}_{x_1} = \hat{\beta}_1$$

Esta última igualdad se cumple sólo si comparamos dos sujetos (hipotéticos) IDENTICOS, excepto porque difieren en UNA UNIDAD en la variable $X_1$

El coeficiente de regresión asociado con la variable `edad` se interpreta como una disminución promedio de 0.0050006 g/cm$^2$ en `dmo` por cada incremento de un año en edad, en mujeres cuya clasificación en `thr` es la misma, es decir, ambas poblaciones de mujeres hipotéticas que usan o no terapia hormonal de reemplazo, pertenecen al mismo grupo del `imc` politómico, y tienen la misma ingesta de calcio y actividad física. De manera similar, se interpreta el coeficiente asociado a la ingesta de calcio. Sin embargo, hay que notar que las unidades de la variable `calcio` son miligramos por día. De ahí que el coeficiente de regresión sea el incremento promedio en `dmo` por cada incremento de un miligramo diario en la ingesta de calcio. Sin embargo, una diferencia de un miligramo diario en calcio total no es nutricionalmente relevante; un cambio de mayor interés es, por ejemplo, 100 mg de calcio diarios. Entonces, el cambio promedio en densidad mineral ósea sería de 0.00104 g/cm$^2$ por cada 100 mg diarios de calcio total en la dieta. El cambio resulta de multiplicar 100 por el coeficiente de regresión asociado con la variable `calcio`.

Con el comando **`lincom`**, pueden obtenerse estimadores de combinaciones lineales de los coeficientes de regresión, así como el intervalo de confianza correspondiente, como en el siguiente ejemplo, en el que se obtiene el cambio promedio en la densidad mineral ósea por cada incremento de 100 mg de calcio total diario en la dieta, con su correspondiente intervalo de confianza:

```
lincom 100*calcio

 ( 1)  100.0 calcio = 0.0
------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0010394   .0004336     2.40   0.017     .0001871    .0018916
------------------------------------------------------------------------------
```

La diferencia en decimales entre multiplicar por 100 el coeficiente de regresión original y los extremos del intervalo de confianza, con el que se obtuvo el estimador por medio del comando **lincom**, sólo se debe al redondeo.

## CUADRO DE ANÁLISIS DE LA VARIANZA

Como en el modelo de regresión lineal simple, los resultados del análisis de la varianza se resumen en el cuadro I. El objetivo es descomponer la variabilidad de *Y* en dos partes: la variabilidad explicada por la regresión y la variabilidad que queda sin explicación y que se expresa por medio de la variabilidad de los residuos. En otras palabras, en regresión lineal múltiple también se cumple que:



$$\underbrace{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}_{\text{variación total}} = \underbrace{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2}_{\text{variación explicada por la regresión}} + \underbrace{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}_{\text{variación de los residuos (no explicada)}} \qquad (1)$$

Para evaluar si la contribución del modelo a la explicación de la variabilidad total del fenómeno es significativa en relación al modelo nulo (modelo que no contiene ninguna variable independiente), se realiza una prueba estadística F. La idea de esta prueba es comparar la variabilidad de *Y* explicada por el modelo con la variabilidad residual a través de un cociente; estas cantidades se encuentran en la tabla de análisis de la varianza en la columna correspondiente a los cuadrados medios [las sumas de cuadrados presentadas en la expresión (1) divididas entre sus respectivos grados de libertad (*k* y *n-k-1*, respectivamente)] que son valores observados de una distribución $\chi^2$. El cociente de estas dos $\chi^2$ genera una estadística de prueba cuya distribución probabilística es una F con *k* y *n-k-1* grados de libertad.

A esta prueba se le conoce como prueba global de la regresión lineal múltiple, ya que compara el modelo propuesto (variabilidad explicada por el modelo) con aquél que se hubiera generado si no se hubiera incluido ninguna variable independiente. En notación matemática se expresa como:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1: \text{por lo menos una } \beta_j \neq 0, \quad j = 1, 2, \ldots, k$$

**Cuadro I Análisis de varianza para regresión lineal múltiple**

| *Fuente de variación* | *Sumas de cuadrados* | *Grados de libertad* | *Cuadrados medios* (CM) | $F_{observada}$ | *Valor* p |
|---|---|---|---|---|---|
| Modelo (regresión) | $SCR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ | k | $CMR = \frac{1}{k}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ | $\frac{CMR}{CME}$ | $P[F_{k,\ n\text{-}k\text{-}1} > F_{observada}]$ |
| Error (residuos) | $SCE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ | $n$-$k$-1 | $CME = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ | | |
| Total | $SCT = \sum_{i=1}^{n}(y_i - \bar{y})^2$ | $n$-1 | | | |

Si se rechaza la hipótesis nula, entonces se concluye que, por lo menos, una variable independiente $X_j$, $j = 1, 2,\ldots, k$ proporciona información significativa para explicar la variabilidad de Y. Si no se rechaza la hipótesis nula, entonces se tiene evidencia de que el conjunto de variables independientes $X_1, X_2, \ldots, X_k$ no ayuda a explicar una cantidad significativa de la variabilidad de $Y$, o bien que la verdadera asociación entre las variables independientes y la de respuesta $Y$ no es lineal.

Hay que notar en el modelo de regresión, generado previamente, que sólo se utiliza información de 430 mujeres, debido a que no todas tienen medición de ingesta diaria de calcio (existen cuatro datos faltantes, por lo que $n = 430$). La prueba global $F$ tiene 6 y 423 grados de libertad porque hay $k = 6$ "variables independientes" en el modelo (`thr`, `edad`, `imc2`, `imc3`, `calcio` y `actfis`) y 430 observaciones (430-6-1 = 423). En realidad, en nuestro marco conceptual sabemos que se trata sólo de cinco variables independientes; pero se contabilizan las dos variables indicadoras que se generaron para caracterizar a `imc`. La prueba global del modelo es altamente significativa, lo que significa que, por lo menos, alguna de las variables independientes (`thr`, `edad`, `imc` politómica, `calcio` y `actfis`) contribuye a explicar la variabilidad de la densidad mineral ósea.



## INFERENCIAS SOBRE LOS COEFICIENTES DE REGRESIÓN

Como se mencionó anteriormente, para poder hacer inferencias sobre los coeficientes de regresión, es necesario, adicionalmente, el supuesto de normalidad de los errores. Bajo estos supuestos, los coeficientes de regresión estimados tienen distribuciones normales con medias iguales a los coeficientes de regresión y varianzas que dependen del parámetro desconocido $\sigma^2$ (varianza del error).

### Intervalos de confianza

Para calcular los intervalos de confianza de los coeficientes de regresión, es necesario primero estimar $\sigma^2$ y después calcular sus errores estándar (*EE*). Una consecuencia de estimar $\sigma^2$ es que la distribución de los coeficientes de regresión es ahora una *t* de Student, y los intervalos de confianza están dados por las siguientes expresiones:

$$\hat{\beta}_j \pm t_{n-k-1}^{1-\alpha/2} \times \widehat{EE}\left(\hat{\beta}_j\right) \qquad j = 0, 1, \ldots, k$$

donde $t_{n-k-1}^{1-\alpha/2}$ es el cuantil $1-\alpha/2$ de una distribución *t* con *n-k-1* grados de libertad.

Cuando el tamaño de muestra, *n*, es "grande", entonces la distribución *t* se aproxima a una distribución normal y, por lo tanto, puede calcularse el intervalo de confianza utilizando la distribución normal en lugar de la *t*. En la práctica, cualquier software estadístico proporciona los errores estándar de los estimadores, así como los intervalos de confianza correspondientes, por lo que no es necesario calcularlos.

Para el ejemplo anterior podemos ver en la salida de STATA la columna correspondiente a los intervalos de confianza asociados con los coeficientes de regresión. Al reportar modelos de regresión, se sugiere presentar los intervalos de confianza y no sólo los valores *p*, porque los intervalos de confianza nos dan una idea de la precisión del estimador. A modo de ejemplo, obsérvese que si se conoce el coeficiente de regresión de la variable `thr` y su error estándar, puede calcularse el intervalo de confianza de la siguiente manera:



$$\hat{\beta}_{thr} \pm t_{423}^{1-0.05/2} \times \widehat{EE}(\hat{\beta}_{thr}) = 0.0106175 \pm 1.96 \times 0.0072816$$

y, por lo tanto, el intervalo de confianza de 95% es $(-0.0036951, 0.02493)$.

### Prueba de hipótesis para el coeficiente de regresión $\beta_j$

Así como en regresión lineal simple era muy importante determinar si el coeficiente asociado con la única variable predictora del modelo era significativamente diferente de cero, en el modelo de regresión lineal múltiple también es de gran interés saber si cada uno de los coeficientes de regresión asociados con las variables independientes son significativamente diferentes de cero, es decir,

$$H_0: \beta_j = 0 \ \ vs. \ \ H_1: \beta_j \neq 0, \qquad j = 1, 2, \ldots, k$$

Los diferentes paquetes estadísticos reportan el valor de la estadística de prueba y el valor *p* al lado del coeficiente de regresión respectivo. La conclusión que se obtiene al rechazar la hipótesis nula es que $X_j$ proporciona información significativa para predecir *Y*. Por otro lado, si no hay evidencia estadística para rechazar la hipótesis nula, entonces se concluye que $X_j$ no contribuye en la predicción de *Y*. Al no poder rechazar la hipótesis nula, también puede sospecharse que la verdadera asociación entre $X_j$ y *Y* no es lineal.

Retomando el ejemplo anterior y la salida correspondiente, se observa que el valor *p* asociado al coeficiente de regresión de la variable `thr` es 0.146, lo que implica que no hay una diferencia significativa (para un nivel de significancia de 0.05) entre las medias de `dmo` para las mujeres que recibieron `thr` con las que no la recibieron, ajustando por edad, índice de masa corporal

politómica, ingesta de calcio y actividad física. Todos los demás coeficientes son altamente significativos e indican que cada coeficiente contribuye a predecir la `dmo` promedio en la presencia de las demás variables. Para la variable actividad física (`actfis`), el valor *p* (0.058) es marginal para un nivel de significancia de 0.05. Entiéndase por marginal que el valor *p* no es estrictamente significativo para este nivel de significancia; pero que está muy cerca de serlo. La decisión de considerar esta asociación significativa o no, concierne directamente al investigador, pues la diferencia con el nivel de significancia convencional (0.05) podría deberse a la variabilidad propia de la muestra de estudio.

**Prueba de hipótesis para la ordenada al origen $\beta_0$**

A diferencia de las pruebas de hipótesis de los coeficientes de regresión, la prueba de hipótesis para la ordenada al origen es de muy poco interés en la mayoría de las aplicaciones. La razón es que la ordenada al origen se interpreta como el valor promedio de *Y*, cuando todas las variables independientes $X_1, X_2,\ldots, X_k$ valen cero, sin importar si este valor está dentro del rango de posibles valores de cada una de las variables. En otras palabras, matemáticamente se requiere estimar la ordenada al origen para poder determinar el modelo de regresión; pero es importante tener presente que no necesariamente tiene interpretación plausible.

## COEFICIENTE DE DETERMINACIÓN



Así como en regresión lineal simple, otro criterio de evaluación para el modelo de regresión lineal múltiple es el coeficiente de determinación, que está dado por la siguiente ecuación:

$$R^2 = \frac{SCT - SCE}{SCT} = \frac{SCR}{SCT}$$

El coeficiente de determinación estima la proporción de la variabilidad de los valores observados de *Y*, que se explica por el modelo de regresión lineal propuesto. Como es una proporción, el rango de posibles valores de $R^2$ es el intervalo [0, 1]; pero en la práctica se expresa comúnmente como un porcentaje. Así, un valor cercano a uno indica que una gran proporción de la variabilidad de *Y* se explica por las variables $X_1, X_2,\ldots, X_k$; mientras que un valor cercano a cero indica que este conjunto de variables ayuda poco a explicar la variabilidad de *Y*. Es importante aclarar que el coeficiente de determinación no es una medida del ajuste del modelo. El supuesto de linealidad se verifica utilizando los residuos del modelo.

Un problema del coeficiente de determinación $R^2$ es que siempre aumenta al introducir nuevas variables en el modelo, independientemente de si su contribución en el ajuste es importante o no. Como en la práctica se prefiere un modelo parsimonioso (es decir, explicar la mayor variabilidad de *Y*, pero con el menor número de parámetros posibles), entonces, con frecuencia, se utiliza una estadística que penaliza el coeficiente de determinación por el número de variables incluidas en el modelo. A esta estadística se le conoce como $R^2$ ajustada y está dada por la siguiente ecuación:

$$R^2_{\text{ajustada}} = R^2 - \frac{k}{n-k-1}\left(1-R^2\right)$$

En la práctica interesan modelos en los que las estadísticas $R^2$ y $R^2_{\text{ajustada}}$ sean lo más parecidas posible, ya que esto sugiere que el número de variables predictoras incluidas en el modelo es adecuado. Cuando esto no sucede, hay variables cuya contribución para explicar la respuesta es prácticamente nula, ya sea porque no aportan información relevante o porque otras variables independientes explican una pieza de información equivalente (variables redundantes).

Retomando el ejemplo anterior y la salida correspondiente, se observa que la $R^2$ es 0.3208 y la $R^2_{\text{ajustada}}$ es 0.3112. Esto indica que poco más de 30% de la variabilidad de `dmo` puede explicarse por las variables: `thr`, `edad`, `imc` politómica, ingesta de calcio y actividad física. La semejanza entre los valores de $R^2$ y $R^2_{\text{ajustada}}$ sugiere que no es un modelo que incluya variables redundantes. La magnitud de $R^2$ es importante, dependiendo del uso del modelo; es especialmente importante cuando quiere utilizarse el modelo para hacer predicciones, situación en la que se busca obtener valores de $R^2$ lo más grande posible.

## EVALUACIÓN ESTADÍSTICA DE LA CONFUSIÓN

Cuando el objetivo de la regresión lineal múltiple sea establecer la relación de una o más variables independientes con la variable dependiente, hay que tener muy en cuenta dos conceptos metodológicos relevantes: la interacción y la confusión. La interacción es la situación en la que la relación de interés difiere de acuerdo con los niveles de una tercera variable involucrada. Para evaluar una interacción, puede emplearse una prueba estadística. Sin embargo, el sustento más fuerte para discutir una interacción es su relevancia clínica y epidemiológica.



Por otro lado, desde el punto de vista estadístico, se dice que se tiene evidencia de confusión cuando la estimación puntual de la relación de interés cambia al ajustar o no en el modelo por la potencial variable confusora. En la práctica, la confusión se evalúa mediante la comparación del estimador crudo de una asociación (que ignora la variable confusora) con el estimador ajustado de la asociación (que controla por la variable confusora). La confusión no se evalúa con una prueba estadística. La importancia relativa del cambio en el estimador de interés se determina subjetivamente a partir del conocimiento del investigador en la disciplina en cuestión.

El estimador crudo de la asociación entre `thr` y `dmo` es 0.0108099 g/cm$^2$, que se obtiene a partir de la siguiente salida (misma que se obtuvo el capítulo de regresión lineal simple), mientras que el estimador de la asociación entre `thr` y `dmo` ajustado por edad, `imc` politómica, ingesta de calcio y actividad física es 0.0106175 g/cm$^2$:

```
regress dmo thr

      Source |       SS       df       MS              Number of obs =     434
-------------+------------------------------           F(  1,   432) =    1.57
       Model |  .008865901     1  .008865901            Prob > F      =  0.2104
    Residual |  2.43448593   432  .005635384            R-squared     =  0.0036
-------------+------------------------------           Adj R-squared =  0.0013
```

```
       Total |  2.44335184   433  .005642845           Root MSE      =  .07507

------------------------------------------------------------------------------
         dmo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .0108099   .0086183     1.25   0.210    -.0061292    .0277491
       _cons |   .3859554   .0040954    94.24   0.000       .377906    .3940047
------------------------------------------------------------------------------
```

Entonces puede decirse que no hay una diferencia importante entre estos dos estimadores y, por lo tanto, en este estudio el efecto de `thr` sobre `dmo` no está confundido por las variables edad, `imc` politómica, ingesta de calcio y actividad física. Cabe señalar que esta aseveración se hace desde un criterio meramente estadístico; sin embargo, debería hacerse una reflexión más profunda de su papel como potenciales confusoras desde el marco epidemiológico, es decir, discutiendo a profundidad su asociación con la variable de respuesta y con la exposición. La decisión de mantener estas variables en el modelo, independientemente de su clasificación como confusoras o no, atiende al hecho de que su influencia sobre la variable de estudio, densidad mineral ósea, ha sido documentada.[5] Una variable puede permanecer en un modelo propuesto por tres razones: 1) por considerarse una variable confusora; 2) por ser una covariable de interés, que aunque no pueda considerarse epidemiológicamente como una variable confusora, tiene una asociación con la variable de estudio previamente documentada, y, finalmente, 3) por su significancia estadística.



## DIAGNÓSTICO DEL MODELO

Se conoce como diagnóstico del modelo a las tres técnicas que se utilizan para verificar que los supuestos teóricos se cumplan en el modelo propuesto. Esta parte del análisis estadístico es de suma importancia, ya que, como se mencionó en el capítulo anterior, los supuestos del modelo son una especie de "requisitos" que deben de cumplir los datos para que sean válidas las inferencias que se realicen a partir del modelo. El paquete estadístico genera el modelo independientemente de que se cumplan estos "requisitos" y es responsabilidad del investigador realizar el diagnóstico del mismo, para tener mayor seguridad de que los resultados que se estén generando son confiables.

Dichas técnicas son: 1) análisis de residuos, 2) puntos extremos y análisis de influencia y 3) estudio de la multicolinealidad. Este diagnóstico se lleva a cabo una vez que ya se tiene un modelo ajustado satisfactorio. Nótese, a modo de recapitulación, que para modelar un fenómeno por medio del modelo de regresión múltiple, se procede de la siguiente manera:

1. Asumir que una relación lineal es un modelo apropiado;

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \quad \varepsilon_i \sim N(0,\sigma^2)\ i = 1, 2, \ldots, n \ \text{ independientes}$$

2. Encontrar y estimar el mejor modelo de regresión lineal (es decir, estimar los coeficientes de regresión). Se busca el modelo más parsimonioso, por medio de la selección de las variables independientes que realmente contribuyan a la explicación de la variable de respuesta; aquéllas que puedan confundir el efecto de la exposición de interés, y aquéllas cuya relevancia se sustente en el marco conceptual del problema.
3. Hacer el diagnóstico del modelo.

**Análisis de residuos**

Este análisis se utiliza para verificar si se cumplen los supuestos estadísticos que se asumen en la generación del modelo propuesto; en particular, los de normalidad, homoscedasticidad y linealidad. Recuérdese que los dos primeros fueron impuestos específicamente para el término de error $\varepsilon_i$ considerado aleatorio. Puesto que este término no es observable directamente de los datos, los supuestos se validarán por medio de los residuos que genera el modelo y que son los mejores estimadores de los errores. Los residuos no son más que la diferencia entre el valor observado de la variable de respuesta y el valor predicho por el modelo: $r_i = y_i - \hat{y}_i$. Estos residuos crudos tienen algunas desventajas técnicas para realizar el diagnóstico, por lo que se han propuesto varios tipos de residuos derivados a partir de ellos: los estandarizados, los estudentizados y los Jacknife. A estos últimos también se les conoce como residuos PRESS.[3] A continuación se presenta una descripción de los más relevantes.

Los residuos crudos ($r_i = y_i - \hat{y}_i$) dependen de las unidades de la variable dependiente. De ahí que no se tenga un criterio homogéneo para decidir si un residuo puede o no considerarse lo suficientemente grande para que se le tome en cuenta como un valor extremo (*outlier* en inglés). Una forma de solucionar este problema es dividir los residuos entre el estimador de la desviación estándar del error ($S = \hat{\sigma}$), para generar lo que se conoce como residuos estandarizados. Éstos tienen la ventaja, sobre los residuos crudos, de que no dependen de las unidades de medición y su interpretación se da en términos de desviaciones estándar. Ambos comparten la desventaja técnica de estar correlacionados entre sí. Por último, los residuos estudentizados-Jacknife se han propuesto como una modificación a ambos tipos de residuos, ya que cuentan con la ventaja adicional de atenuar los problemas de dependencia y se consideran los más recomendables para realizar el diagnóstico.

El comando **`predict`**, en STATA, genera distintas variables de interés, dependiendo de la opción que se especifique de acuerdo con el último modelo que se haya corrido. Por ejemplo, para calcular los valores predichos $\hat{y} = \hat{\text{dmo}}$, se especifica la opción **`xb`** y se proporciona el nombre que se dará a la variable que contendrá los valores predichos. En el ejemplo, a esta variable se le llama `y_gorro`. En la siguiente salida, puede verse que los valores predichos a veces están por arriba y a veces por abajo del valor observado, lo que implica que habrá residuos negativos y positivos, respectivamente. Para una expresión gráfica de esto, se recomienda ver la figura 5 del capítulo de regresión lineal simple.

```
predict y_gorro, xb

list dmo y_gorro in 1/5
     +-----------------+
     |  dmo    y_gorro |
     |-----------------|
  1. | .431   .3700744 |
  2. | .369   .3780156 |
  3. | .417   .4257628 |
  4. | .405   .4009440 |
  5. | .550   .4130634 |
     +-----------------+
```

Con el comando **predict** también pueden generarse los residuos crudos y los Jacknife. En la siguiente salida se muestra cómo generar estos dos tipos de residuos, un listado de las primeras cinco observaciones en la base de datos y las estadísticas descriptivas de los residuos:

```
predict residuos, residuals
(4 missing values generated)

predict res_jack, rstudent
(4 missing values generated)
```



Nótese que se generan cuatro valores faltantes ($n = 430$ en lugar de las 434 mujeres que conforman la base de datos), debido a que no puede estimar el valor predicho de las cuatro mujeres que no tienen medición de la ingesta diaria de calcio (`calcio`):

```
list dmo y_gorro ths edad  imc2 imc3 calcio actfis residuos res_jack in 1/5

   +-------------------------------------------------------------------------+
   |  dmo    y_gorro  ths edad imc2 imc3 calcio   actfis   residuos  res_jack  |
   |-------------------------------------------------------------------------|
1. | .431   .3700744   1   52   0    0    730.93      0    .0609256   .9944581 |
2. | .369   .3780156   0   49   0    0   1073.14      0   -.0090156  -.1462606 |
3. | .417   .4257628   0   48   0    1    811.09  1.3125  -.0087628  -.1419506 |
4. | .405   .400944    0   52   1    0    616.82 16.25     .004056    .0657172 |
5. |  .55   .4130634   0   48   1    0    772.72      0    .1369366  2.230943  |
   +-------------------------------------------------------------------------+

summ residuos res_jack

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    residuos |     430   -3.36e-11    .0615278  -.2285229   .1726197
    res_jack |     430     .000243    1.003733  -3.770745   2.827965
```

Entonces, por ejemplo, pueden confirmarse los cálculos de los valores predichos y de los residuos crudos para la primera mujer listada:

$$\widehat{dmo} = 0.6118931 + 0.0106175 \times 1 - 0.0050006 \times 52 + 0.0331696 \times 0 + 0.0447026 \times 0 + 0.0000104 \times 730.93 + 0.0005848 \times 0 = 0.37008$$

$$\textit{RESIDUO CRUDO} = dmo - \widehat{dmo} = 0.431 - 0.37008 = 0.06092$$

De las estadísticas descriptivas de los residuos, se confirma que la media de los residuos crudos es cero (matemáticamente lo es; pero debido al redondeo en las computadoras, sale un número muy pequeño) y que toman valores en las mismas unidades que la variable dependiente dmo. Mientras que los residuos Jacknife tienen una media casi igual a cero (pero no es cero) y desviación estándar prácticamente igual a uno; es decir, las unidades de medición están dadas en desviaciones estándar y no en la escala original. La forma más común de evaluar los supuestos es por medio de diferentes gráficas. En algunas de ellas se busca que no haya patrones *evidentes* (abanicos, herraduras, grupos separados de datos, tendencias de ningún tipo, etc.).

Para evaluar el supuesto de normalidad puede hacerse un histograma de los residuos, un diagrama de caja (también conocido como *boxplot*), o una gráfica de los cuantiles de la variable de interés contra los cuantiles de una distribución normal (también conocida como diagrama cuantil-cuantil, o *Q-Q* plot). Otras estadísticas de comparación pueden ser el sesgo y la kurtosis, que para la distribución normal valen cero y tres respectivamente.



Bajo el supuesto de normalidad, en el histograma se esperaría ver una curva en forma de campana; en el diagrama de caja, se esperaría una caja simétrica respecto a la media y a la mediana, y en el diagrama cuantil-cuantil se esperaría que los datos estén lo más cercanos posible a la recta de 45° (ya que ésta refleja los cuantiles de la distribución de una variable con distribución idéntica a una normal). De estos tres instrumentos gráficos, se considera que el más recomendable es el diagrama cuantil-cuantil, ya que analiza los residuos de manera individual, sin agrupar. Al definir las barras que generarán el histograma, se pierde el detalle de lo que sucede en el interior de cada categoría, y si se aumenta el número de barras de manera excesiva, se corre el riesgo de perder la forma global de la distribución. El diagrama de cajas, en contraste, presenta de una manera muy completa la distribución de los residuos, y hace especial énfasis en lo que se refiere a la simetría de la distribución; sin embargo, se pierde el detalle de lo que ocurre en el interior de la caja misma. Véanse las gráficas descriptivas (histograma y de caja) de los residuos Jacknife (figura 2).

Para probar normalidad en los errores, algunos autores sugieren[1] utilizar la prueba de bondad de ajuste de Shapiro-Wilk para tener un criterio cuantitativo que apoye su decisión ($H_o$: los residuos tienen una distribución normal); sin embargo, el uso de esta prueba debe hacerse con cierta cautela, debido a que el supuesto de independencia de las observaciones de esta prueba se viola cuando se aplica sobre los residuos; incluso si se utilizan los residuos Jacknife, donde la dependencia está atenuada, pero que no deja de existir.

A continuación se muestra la salida de STATA donde se evalúa el supuesto de normalidad en el modelo de regresión lineal múltiple, ejemplificado a lo largo de esta sección. De acuerdo con la prueba de Shapiro-Wilk hay evidencia estadística de que la distribución de los residuos es normal, ya que no se rechaza la hipótesis nula (valor $p = 0.31036$). En el diagrama cuantil-cuantil

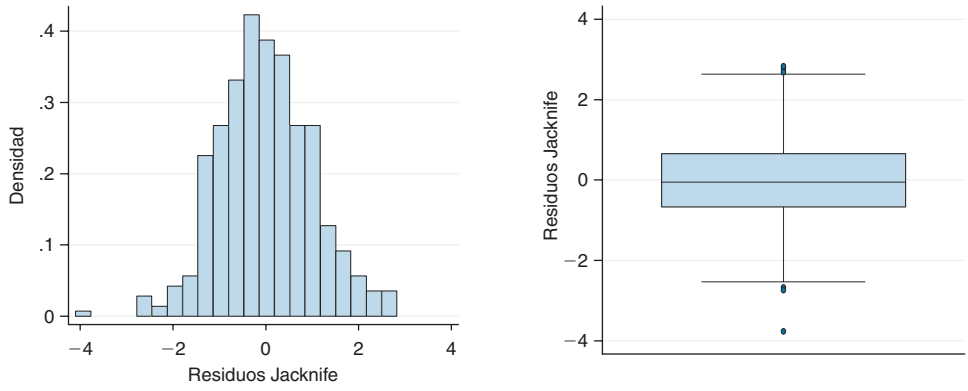



**Figura 2** **Histograma y gráfica de caja de los residuos**

de la figura 3 se observa que, con excepción de algunos puntos en las colas, la distribución de los residuos tiene un comportamiento muy semejante al de una distribución normal porque la nube de puntos está muy cerca de la recta diagonal. Los residuos tienen una distribución ligeramente sesgada a la derecha (sesgo de 0.05) y es un poco más alargada que la normal (kurtosis de 3.33); sin embargo estos valores son muy cercanos a los de una distribución normal. Además, sólo el 5.6 y 8.8% de los residuos Jacknife exceden en valor absoluto a 1.96 y 1.645, respectivamente; estos porcentajes son muy similares a 5% y 10% esperados, en una distribución normal.



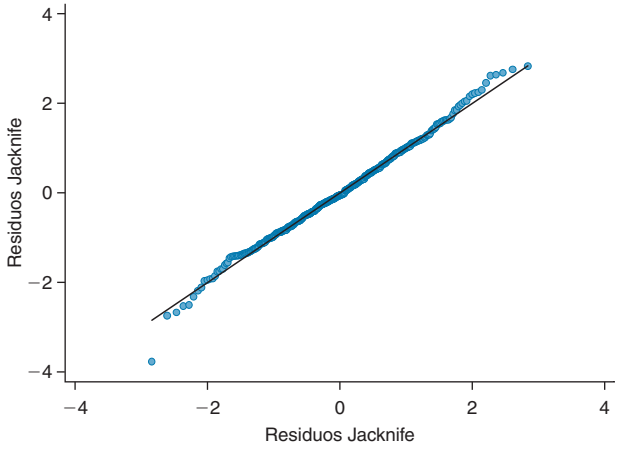



**Figura 3** **Diagrama cuantil-cuantil de los residuos**

```
swilk res_jack

                    Shapiro-Wilk W test for normal data
    Variable |    Obs        W          V          z      Prob>z
-------------+-------------------------------------------------
    res_jack |    430    0.99581      1.230      0.495   0.31036

summ res_jack, d

                    Studentized residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -2.501476      -3.770745
 5%    -1.434202      -2.746166
10%    -1.248825      -2.674971       Obs                 430
25%    -.6728688       -2.53313       Sum of Wgt.         430

50%    -.0557179                      Mean            .000243
                        Largest       Std. Dev.       1.003733
75%     .6521232        2.63366
90%     1.223501        2.681518      Variance         1.00748
95%     1.623226        2.757436      Skewness        .0517896
99%     2.618459        2.827965      Kurtosis        3.335767
```



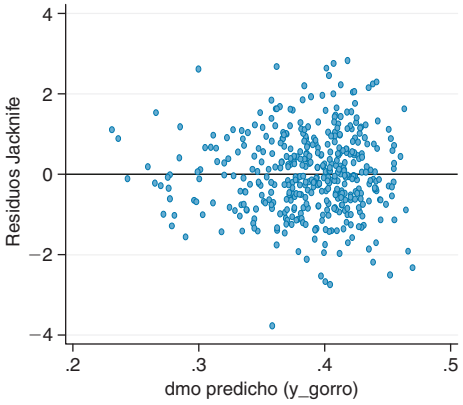
Para evaluar el supuesto de homoscedasticidad, se grafican los residuos Jacknife contra los valores predichos de la variable de respuesta (figura 4). Una evidencia de que este supuesto no se cumple es cuando se observa algún patrón; el más común se presenta en forma de abanico y refiere una tendencia de la varianza a incrementarse conforme el valor de la variable de respuesta se acerca a alguno de sus extremos. STATA tiene el comando **`rvfplot`** (*residual versus fitted plot*) para hacer esta gráfica, utilizando los residuos crudos, pero no lo tiene implementado con los residuos Jacknife. Puesto que éstos no son más que un reescalamiento de los residuos crudos con la característica adicional de abordar el problema de independencia —que aquí no interfiere—, no se esperaría que la gráfica fuera muy diferente a la generada con los residuos Jacknife.

Teóricamente, la mitad de los residuos, aproximadamente, deben estar dispersos por arriba y la mitad por debajo de la recta $Y = 0$. En esta gráfica, no se observa ningún patrón. Si el modelo es adecuado, se esperaría que hubiera pocos residuos extremos (muy chicos o muy grandes), por lo que en los extremos del rango de los valores estimados del eje $X$, habría menos puntos y, como consecuencia, parece verse menos variabilidad; pero este efecto es producto de que se tienen menos puntos en la gráfica.

Debido a que la interpretación de las gráficas es subjetiva, en ocasiones, es necesario auxiliarse de alguna prueba estadística para complementar cuantitativamente la apreciación de la gráfica. STATA tiene implementada la prueba estadística de homoscedasticidad de Cook-Weisberg[6] con el comando **`hettest`**. La hipótesis nula de esta prueba es que la varianza es constante. En el ejemplo, la prueba no es significativa (valor $p = 0.1280$); es decir, hay evidencia estadística de que la varianza es constante.

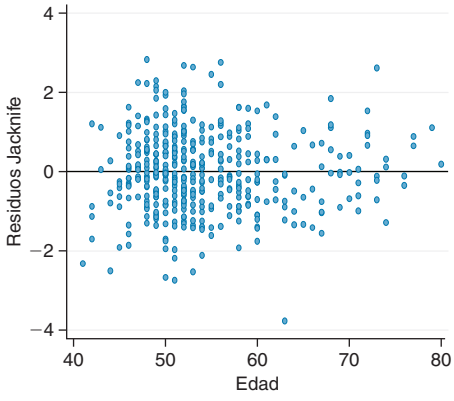
SUPUESTO DE HOMOSCEDASTICIDAD

```
scatter res_jack y_gorro, yline(0)
```



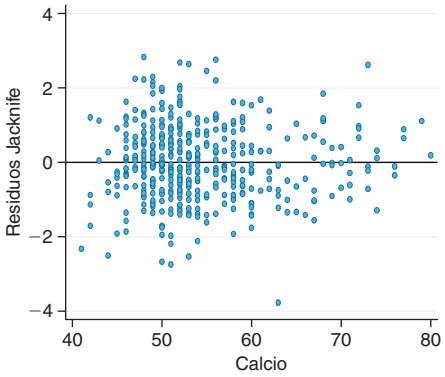

SUPUESTO DE LINEALIDAD

```
scatter res_jack edad, yline(0)
```




SUPUESTO DE LINEALIDAD

```
scatter res_jack calcio, yline(0)
```




SUPUESTO DE LINEALIDAD

```
scatter res_jack actfis, yline(0)
```

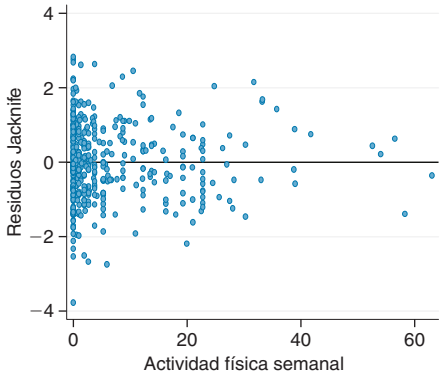



**Figura 4** **Evaluación gráfica de los supuestos de homoscedasticidad y linealidad**

```
hettest

Cook-Weisberg test for heteroskedasticity using fitted values of dmo
     Ho: Constant variance
         chi2(1)      =      2.32
         Prob > chi2  =      0.1280
```

El supuesto de linealidad puede verificarse por medio de la gráfica de los residuos Jacknife *versus* cada una de las variables independientes (figura 4). STATA tiene el comando **`rvpplot`** (residual versus predictor plot) para hacer esta última gráfica, usando los residuos crudos en lugar de los Jacknife. Si el supuesto se cumple, no debería detectarse ningún patrón en la gráfica. En el caso de regresión lineal simple, el supuesto de linealidad también puede verificarse por medio de la misma gráfica utilizada para evaluar el supuesto de homoscedasticidad[1] (residuos Jacknife *versus* los valores predichos de la variable de respuesta). En ninguna de las cuatro gráficas se observa alguna tendencia particular.

Por último, el supuesto de independencia de los errores, generalmente, se verifica por el diseño del estudio, esto es, analizando cualitativamente si existen características de las observaciones que pudieran generar correlación entre ellas. Por ejemplo, si se trata de una encuesta nacional, habría que reflexionar si el hecho de que los sujetos habiten en una misma localidad que comparte condiciones socioeconómicas, ambientales y sanitarias, entre otras, pudiera obtenerse valores más semejantes en la variable dependiente comparados con los que se obtendrían en una muestra aleatoria. En un estudio epidemiológico es importante reflexionar cómo afectaría a la independencia de las observaciones tener miembros de una misma familia en la población de estudio. Finalmente, el caso más evidente de observaciones correlacionadas se presenta cuando se hace un estudio longitudinal con mediciones repetidas en el tiempo sobre un mismo sujeto.

**Puntos extremos y análisis de influencia**



Un valor extremo (*outlier* en inglés) es una observación rara o inusual: es un valor cuyo residuo ($r_i = y_i - \hat{y}_i$) es "mucho" más grande en valor absoluto que el resto (el correspondiente residuo Jacknife es, en valor absoluto, mayor que tres, identificando así, residuos crudos que están a más de tres desviaciones estándar de la media). Los valores extremos pueden ser, aunque no necesariamente, altamente influyentes en el resultado del análisis y por eso merecen atención. Un punto influyente es aquel que tiene un impacto excesivo sobre el ajuste del modelo, en el sentido de que la estimación numérica de los parámetros o las predicciones pueden depender más de este punto que de la mayoría del conjunto de datos. Por ejemplo, en las figuras 5a y 5b se ejemplificó la influencia sobre la pendiente de la recta para dos valores extremos (puntos 1 y 2). En estas gráficas se observan simultáneamente dos modelos de regresión: una recta que incluye al punto extremo y otra que lo excluye. Podemos ver que la recta que excluye al punto extremo 1 ($R_{\text{sin } 1}$) tiene una pendiente muy similar a la recta que no lo excluye ($R_{\text{con } 1}$). Es decir, el punto extremo 1 no influye fuertemente en la pendiente de la recta estimada, mientras que en la figura 5b, el punto extremo 2 sí influye, porque al excluirlo del modelo de regresión ($R_{\text{sin } 2}$), la pendiente de la recta cambia considerablemente respecto a la recta que lo incluye ($R_{\text{con } 2}$).

Una vez que se han detectado los valores extremos, debe evaluarse primero su plausibilidad y después su importancia; es decir, primero se determina si esos valores son plausibles y se descarta que no sean posibles errores de captura o de levantamiento de la información. Aunque se supone que el análisis partió de una base de datos limpia y ampliamente verificada, siempre es posible que el diagnóstico del modelo ayude a detectar valores extremos en el espacio multidimensional que definen todas las variables independientes. Por ejemplo, en un estudio que incluya las variables peso y edad, se parte del supuesto de que, de inicio, se realizó una limpieza de la base de datos donde se verificó la plausibilidad de los datos de cada una de estas variables de manera individual, por lo que un residuo muy grande de un modelo múltiple que incluye-

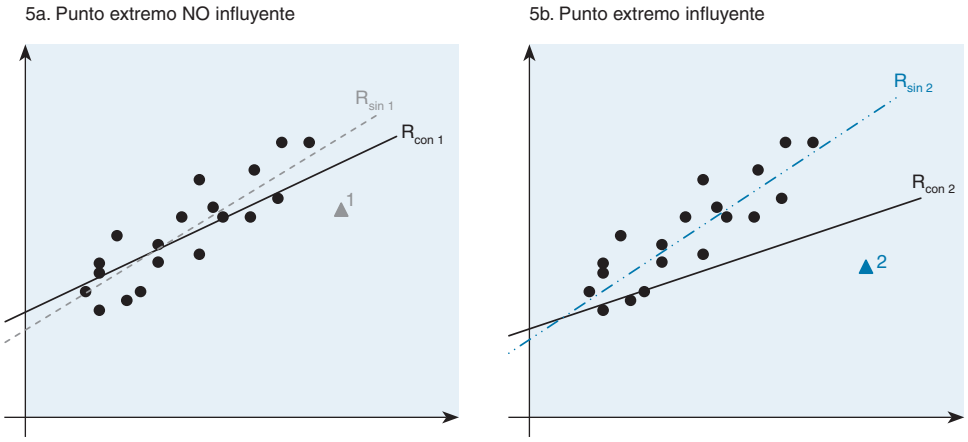



**Figura 5** **Influencia de dos valores extremos sobre la pendiente de la recta**

ra estas dos variables estaría sugiriendo que existe un sujeto, cuya combinación de peso y edad es rara por sí misma y/o en relación con el valor observado de la variable de respuesta que estuviera estudiándose. Una vez que se verificó que los valores extremos sean plausibles, entonces se procede a realizar un análisis de influencia de las estimaciones sobre los coeficientes de regresión estimados. Es importante destacar que son conceptos diferentes: los puntos extremos pueden o no ser puntos influyentes y pueden existir valores influyentes que no se hayan detectado como valores extremos.

Las estadísticas que se utilizan para identificar puntos influyentes se basan en comparaciones entre los estimadores que se obtienen con y sin la presencia del punto en el modelo. Intuitivamente, los puntos influyentes serán aquellos cuya inclusión o exclusión genere diferencias grandes en los estimadores obtenidos; mientras que la inclusión o exclusión de los puntos no influyentes será prácticamente imperceptible en las estimaciones. Para determinar si el valor de alguna estadística de influencia debe considerarse suficientemente grande para llamar la atención hacia alguna observación, es necesario definir puntos de corte que especifiquen este umbral. Es importante señalar que las recomendaciones estadísticas que se dan para estos puntos de corte sólo son lineamientos y deben utilizarse con reserva. El criterio y experiencia del investigador son indispensables para decidir si la magnitud del cambio observada es relevante o no en el contexto del problema. A veces los cambios son importantes en un contexto específico del problema, aun cuando las estadísticas de diagnóstico no rebasen el valor formal del punto de corte.

Ahora bien, aun cuando se detecten puntos estadísticamente influyentes, se recomienda hacer el análisis con y sin estas observaciones para comparar los resultados. Puede ser que ciertos puntos influyan estadísticamente en los estimadores; pero de manera que en la práctica los cambios sean tan minúsculos, que biológica o clínicamente sean irrelevantes. Es muy importante explorar cuidadosamente los puntos influyentes y/o extremos antes de proponer su exclusión, ya que

la exclusión excesiva de observaciones puede introducir un sesgo de selección o, incluso, perder la validez interna y externa del estudio. Además, las conclusiones obtenidas de este análisis deben incluirse en el reporte científico del estudio.

La forma para detectar los puntos extremos es por medio de los residuos Jacknife, y los puntos palanca (*leverage* en inglés). El punto palanca $h_i, i = 1,2,\ldots, n$ mide qué tan extrema es la observación $i$ en relación con el espacio generado únicamente por las variables independientes. De hecho, es una medida de distancia estandarizada entre la observación $(x_{i1}, x_{i2},\ldots, x_{in})$ y el centroide $(\overline{x}_1 = \overline{x}_2,\ldots, \overline{x}_k)$. Entonces, los puntos palanca con valores grandes sugieren que están alejados del espacio generado por las variables independientes. Si el modelo de regresión tiene ordenada al origen, los puntos palanca toman valores entre $1/n$ y 1. Un punto de corte que se utiliza para detectar puntos palanca grandes es $2(k + 1)/n$ ya que corresponde al doble del promedio de los puntos palanca. En STATA, los puntos palanca se obtienen con la opción **`hat`** del comando **`predict`**.

Los valores extremos pueden influir en las estadísticas de resumen del análisis de regresión de las siguientes tres formas:

a) simultáneamente en todos los coeficientes de regresión;
b) sólo en algún coeficiente de regresión $\beta_j$, y
c) en los valores predichos $\hat{y}_i$.



Para detectar los puntos que influyen sobre todos los coeficientes de regresión, puede utilizarse la distancia de Cook $D_i, i = 1, 2,\ldots, n$. Para cada observación de la muestra, la distancia de Cook $D_i$ mide qué tan diferentes son los estimadores de los coeficientes de regresión que se obtendrían con y sin la $i$-ésima observación. Además, puede demostrarse que la distancia de Cook depende tanto del residuo como del punto palanca, por lo que una distancia de Cook grande puede deberse a que la observación es un valor extremo en el espacio de las variables independientes o a que la observación tiene un residuo grande. Si el modelo es correcto, se espera que las distancias de Cook para todas las observaciones sean menores que el punto de corte, que es 1. En caso de que haya observaciones con distancias de Cook mayores a 1, entonces se recomienda hacer el análisis con y sin estas observaciones y comparar los resultados. En STATA, las distancias de Cook $D_i$ se obtienen con la opción **`cooksd`** del comando **`predict`**.

La estadística para detectar puntos extremos que pueden influir en el coeficiente de regresión $\beta_j$ es $\text{dfbeta}_j$ e indica cuánto cambia este coeficiente de regresión, si se omitiera la $i$-ésima observación. Un valor grande de $\text{dfbeta}_j$ indica que la $i$-ésima observación influye sobre el $j$-ésimo coeficiente de regresión. Por otro lado, una estadística para detectar puntos extremos que pueden influir en los valores predichos es dfit y también mide la influencia de la eliminación de la $i$-ésima observación sobre el valor predicho.[3] STATA tiene estas estadísticas implementadas mediante los comandos **`dfbeta`** y **`dfit`**, respectivamente.

**Multicolinealidad**

El problema de multicolinealidad se define como el exceso de correlación entre las variables independientes, lo que puede generar inestabilidad numérica, es decir que los estimadores sean muy diferentes para dos muestras de la misma población. El término colinealidad se refiere a que una variable es exactamente una combinación lineal de las otras variables. Cuando esto sucede, bas-

ta con sacar del modelo a la variable que es redundante porque simplemente no contribuye en la explicación del fenómeno. Aun cuando hay casi colinealidad (dependencia lineal), es posible ajustar el modelo; pero los estimadores son muy sensibles a cambios en los datos, es decir, que si se hubiera obtenido otra muestra de datos, los estimadores podrían ser muy diferentes.

Para encontrar las variables que podrían generar problemas de multicolinealidad en el modelo, una de las estadísticas que puede utilizarse es la del factor de inflación de la varianza (*variance inflation factor, VIF*).

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \ldots, k$$

donde $R_j^2$ es la $R^2$ que se habría obtenido al evaluar la capacidad explicativa que tienen $k$-1 variables independientes sobre la restante $X_j$, la cual juega el papel de variable dependiente en un modelo de regresión auxiliar.

Valores altos del $VIF_j$ son evidencia de que la variable $X_j$ puede explicarse en gran medida por las variables restantes. Una recomendación utilizada frecuentemente por los estadísticos es considerar un $VIF_j$ grande, si éste es mayor a 10, lo cual se obtiene cuando se alcanza una $R_j^2 > 0.9$. Para una mayor comprensión de la multicolinealidad, ver posibles formas de evitarla y solucionarla, véanse libros especializados en regresión lineal.[1,3] En STATA, mediante el comando **vif** puede obtenerse simultáneamente el factor de inflación de la varianza $VIF_j$ para cada variable independiente.

Por último, se recomienda que cuando quiera generarse un modelo de regresión lineal, no sólo se limite a la información presentada en estos dos capítulos, sino que se profundice en otros temas, como son los criterios para generar un modelo y la generación de intervalos de predicción, entre otros.



## REFERENCIAS


1. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Applied regression analysis and other multivariable methods. 3a. ed. Pacific Grove, CA: Duxbury Press, 1998.
2. Mardia KV, Kent JT, Bibby JM. Multivariate analysis. Nueva York: Academic Press, 2000:213-216.
3. Montgomery DC, Peck EA. Introducción al análisis de regresión lineal. 1a. ed. México: CECSA, 2002.
4. Draper NR, Smith H. Applied regression analysis. Washington: A Wiley-Interscience Publication, 1998.
5. López-Caudana AE, Téllez-Rojo MM, Hernández-Ávila M, *et al*. Predictors of bone mineral density in Mexican Social Security Institute's female workers in Morelos State, Mexico. Archives of Medical Research 2004;35:172-180.
6. Goldstein R. Cook-Weisberg test of heteroskedasticity. Stata Technical Bulletin Reprints 1992;2:183-184.

# XIV Regresión logística

*Martha María Téllez Rojo Solís*
*Héctor Lamadrid Figueroa*
*Daniela Sotres Álvarez*

## INTRODUCCIÓN

El presente capítulo tiene como objetivo presentar el modelo de regresión logística en el contexto de su aplicación en un estudio epidemiológico. En él se describen las características estadísticas del modelo, haciendo especial énfasis en la interpretación y uso adecuado del mismo. Hasta ahora, el material presentado en los capítulos previos sobre regresión lineal simple y múltiple, se refirió específicamente a problemas en los que se pretendió estimar la asociación entre un conjunto de variables independientes o predictoras y una variable de respuesta continua. Ahora, se iniciará un recorrido por la estimación de este tipo de asociaciones; pero cuando la variable de respuesta es de tipo dicotómico, es decir, cuando sólo puede tomar dos valores, los cuales comúnmente se refieren a la presencia o ausencia de una característica, de un diagnóstico o, en general, de cualquier condición de salud.



Se incursionará desde la situación más simple y se retomará el ejemplo del capítulo anterior, en el que se realizó un estudio transversal en 434 mujeres menopáusicas trabajadoras del IMSS, delegación Morelos, con el objetivo de determinar algunos factores asociados a la densidad mineral ósea (`dmo`). A diferencia del capítulo anterior, en el que se trabajó con la `dmo` como variable continua, pensemos ahora que se quieren identificar algunos factores que se asocian con la presencia de osteoporosis en esta población, específicamente, su asociación con el uso de terapia hormonal de reemplazo (`thr`). Para ello, se clasificó este grupo de estudio en dos categorías, con los siguientes criterios médicos: aquellas mujeres cuya `dmo` fue menor a 0.330 g/cm$^2$ y aquellas con valores por arriba de este punto de corte. Las primeras fueron diagnosticadas con osteoporosis y se identificaron con el valor 1 de una variable con este mismo nombre. Las restantes fueron asignadas a la categoría 0, que refleja la ausencia de este diagnóstico. Con esta idea en mente, se quiere conocer la asociación entre el uso de `thr`, variable independiente o exposición, y la variable dependiente o de respuesta, `osteoporosis`. Lo cual se resume en la figura 1:

La variable de respuesta sólo tiene dos categorías: presencia o ausencia de diagnóstico de osteoporosis, las cuales se identifican con los valores 1 y 0 de la variable `osteoporosis` respectivamente. Por otra parte, se tiene una variable independiente que, en este ejemplo, también fue de tipo dicotómico, ya que clasifica a las mujeres que utilizan `thr` con la categoría 1, y aquellas que no la utilizan las identifica con la categoría 0. No obstante, podría tenerse una medida más precisa de esta variable, como saber cuál ha sido el tiempo (p. ej. número de años) que lleva cada participante utilizando este tipo de terapia o, incluso, conocer la cantidad de estrógenos contenida en cada uno de los medicamentos utilizados a través del tiempo. Como puede verse, en estas

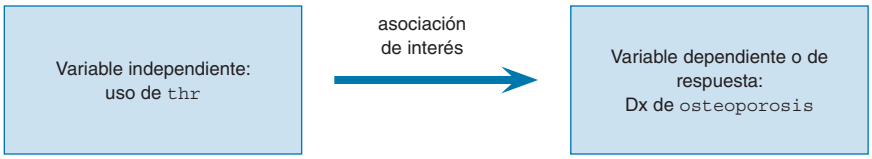



**Figura 1** **Asociación entre terapia hormonal de reemplazo (thr) y osteoporosis**

dos últimas situaciones se estaría buscando refinar la medida de exposición para pasar de tener una variable dicotómica a una discreta (número de años) y, finalmente, a una continua (dosis).

Es importante hacer notar que la restricción del modelo de regresión logística de tener una variable que sólo puede tener dos categorías, se impone únicamente sobre la variable dependiente, que en nuestro caso es `osteoporosis`. Para facilitar la presentación, piénsese en una única variable independiente (exposición) dicotómica; pero más adelante extiéndase esta situación a una variable discreta, a una continua y, finalmente, a un número mayor de variables independientes.

Una variable como `osteoporosis` se conoce en la jerga estadística como una variable Bernoulli con un único parámetro de interés: la probabilidad de que un sujeto presente el evento de estudio. Esta probabilidad se conoce como $P$, y la distribución de la variable se denota de la siguiente manera:



$$\text{osteoporosis} \sim \text{Bernoulli}(p)$$

donde:

$$P[\text{osteoporosis} = 1] = p$$
$$P[\text{osteoporosis} = 0] = 1 - p$$

Si esto se vincula con la parte introductoria del libro, recuérdese que una medida de asociación adecuada para un estudio transversal como el que se describe previamente, con un evento y exposición dicotómica, es la razón o cociente de momios. Un breve resumen estadístico de lo que hay detrás de esta medida sería el siguiente:

Un momio compara la probabilidad de ocurrencia de un evento con la probabilidad de que *no* ocurra, en una población dada, bajo las mismas condiciones. Si un evento ocurre con probabilidad $P$, entonces el momio de ocurrencia de ese evento en la población expuesta al factor de riesgo se define como

$$\text{momio}_{\text{expuesto}=1} = \frac{P\left[\text{evento} = 1 \,|\, \textit{expuesto} = 1\right]}{P\left[\text{evento} = 0 \,|\, \textit{expuesto} = 1\right]}$$

Para calcular el momio estimado en el ejemplo, se resume primero la información disponible en una tabla de $2 \times 2$:

| | | Exposición: uso de thr | | |
|---|---|---|---|---|
| | **Diagnóstico** | 1: Sí | 0: No | **Total** |
| **Evento: osteoporosis** | 1 : dx positivo | $a$ | $b$ | $a + b$ |
| | 0 : dx negativo | $c$ | $d$ | $c + d$ |
| | Total | $a + c$ | $b + d$ | |

entonces, el momio para el grupo que usa `thr` (*i.e.* el grupo expuesto: `thr=1`) sería:

$$\text{momio}_{thr=1} = \frac{P\left[\text{osteoporosis} = 1 \mid thr = 1\right]}{P\left[\text{osteoporosis} = 0 \mid thr = 1\right]} = \frac{\frac{a}{a+c}}{\frac{c}{a+c}} = \frac{a}{c}$$

y el momio correspondiente para el grupo no expuesto:

$$\text{momio}_{thr=0} = \frac{P\left[\text{osteoporosis} = 1 \mid thr = 0\right]}{P\left[\text{osteoporosis} = 0 \mid thr = 0\right]} = \frac{\frac{b}{b+d}}{\frac{d}{b+d}} = \frac{b}{d}$$



*Observaciones relevantes*: Un momio *no* es una probabilidad, sino *un cociente* de probabilidades. Es un número mayor o igual a cero, tan grande como grandes sean las posibilidades de tener un diagnóstico positivo para osteoporosis, en comparación a no tenerlo cuando se restringe la comparación a un grupo de sujetos con un factor (o factores) de riesgo común.

Ahora bien, una alternativa para evaluar qué tanto se asocia el uso de `thr` con el hecho de tener un diagnóstico positivo para osteoporosis sería comparar estos dos momios, ya que contrastan la posibilidad del diagnóstico positivo en las dos situaciones de interés: uso y no uso de `thr`. Esta comparación podría darse por medio de una diferencia de momios, o bien de una razón. En el primer caso, el valor nulo, es decir, la ausencia de asociación entre la exposición y el evento, se vería reflejado con una diferencia de momios igual a 0. En caso de que estos momios se compararan mediante una razón, el valor nulo sería el 1. Ambos casos reflejarían situaciones idénticas, en donde la probabilidad de que un sujeto tenga diagnóstico de osteoporosis en relación a no tenerlo, es muy similar entre los participantes expuestos y no expuestos. Las dos alternativas son igualmente válidas; sin embargo, la razón de momios (*RM*) tiene propiedades estadísticas y epidemiológicas que la harán más atractiva y que se desarrollarán más adelante.

Por lo anterior, se define RM como,

$$RM = \frac{\text{momio}_{\text{expuesto} = 1}}{\text{momio}_{\text{expuesto} = 0}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc} \qquad (1)$$

Una $RM > 1$ será evidencia de que la exposición se asocia con mayores posibilidades de desarrollar la enfermedad, mientras que una $RM < 1$ reflejará la asociación con una exposición protectora. La interpretación numérica de una *RM* se realizará en términos multiplicativos, ya que la forma de comparación usada entre estos momios fue un cociente.[1] Véase qué sucede en este ejemplo y cómo se calcularía la *RM* en el paquete estadístico STATA:

1. Primero, conózcanse las variables y su codificación. Se destacan en "negritas" las palabras que refieren comandos específicos del paquete STATA.

```
desc osteoporosis thr
              storage   display     value
variable name   type    format      label      variable label
------------------------------------------------------------------------------
osteoporosis    float   %9.0g                 1: osteoporosis, 0:no
                                                  osteoporosis
thr             float   %9.0g                 Uso de terapia hormonal 1=si
                                                  0=no
```

2. Por medio de una tabulación cruzada de estas variables, pueden generarse los momios correspondientes.



```
tab osteoporosis thr
        1: |
osteoporos |
  is, 0:no |    Uso de terapia
osteoporos |  hormonal 1=si 0=no
        is |         0          1 |     Total
-----------+----------------------+----------
         0 |       253         83 |       336
         1 |        83         15 |        98
-----------+----------------------+----------
     Total |       336         98 |       434
```

Cabe señalar que el paquete automáticamente ordena las columnas y los renglones de acuerdo con la categoría numérica asignada, por lo que deberá realizarse la adaptación correspondiente a lo antes expuesto.

3. Ahora, realícese el cálculo manual de la *RM*.

$$\text{momio}_{thr=1} = \frac{P\left[\text{osteoporosis} = 1 \mid thr = 1\right]}{P\left[\text{osteoporosis} = 0 \mid thr = 1\right]} = \frac{\frac{15}{98}}{\frac{83}{98}} = 0.181$$

$$\text{momio}_{thr=0} = \frac{P\left[\text{osteoporosis} = 1 \mid thr = 0\right]}{P\left[\text{osteoporosis} = 0 \mid thr = 0\right]} = \frac{\frac{83}{336}}{\frac{253}{336}} = 0.328$$

$$RM = \frac{\text{momio}_{thr=1}}{\text{momio}_{thr=0}} = \frac{0.181}{0.328} = 0.552$$

4. Ahora, automáticamente.

```
tabodds osteoporosis thr, or

------------------------------------------------------------------------------
         thr |  Odds Ratio       chi2       P>chi2     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           0 |    1.000000          .          .               .           .
           1 |    0.550878       3.82       0.0506      0.300347    1.010388
------------------------------------------------------------------------------

Test of homogeneity (equal odds): chi2(1)   =     3.82
                                  Pr>chi2   =   0.0506
Score test for trend of odds:     chi2(1)   =     3.82
                                  Pr>chi2   =   0.0506
```

5. *Interpretación*. El momio de presentar un diagnóstico positivo de osteoporosis en el grupo que utiliza `thr` es prácticamente la mitad del momio correspondiente para el grupo que no utiliza la terapia hormonal (equivalentemente, el momio del grupo que no utiliza `thr` es prácticamente el doble de aquél para el grupo de mujeres que sí utiliza `thr`). Como la $RM < 1$, puede interpretarse que la exposición a `thr` es protectora para el desarrollo de osteoporosis.

## EL MODELO DE REGRESIÓN LOGÍSTICA

Esta sección estará dedicada a responder las siguientes preguntas: ¿cómo se relaciona lo anteriormente expuesto con el modelo de regresión logística?; ¿qué ventajas tiene incorporar el enfoque del modelo de regresión logística sobre el conocimiento que ya se tenía de cursos básicos de epidemiología?; ¿cuáles son las generalizaciones que este enfoque permite realizar para que "valga la pena" aprenderlo?; ¿cómo se vinculan estos nuevos conceptos con lo aprendido en el capítulo anterior sobre el modelo de regresión lineal?

En primer lugar, se dará una síntesis de las semejanzas y diferencias que hasta el momento pueden mencionarse entre lo expuesto y un modelo de regresión lineal simple.

**Cuadro I Comparación entre los modelos de regresión lineal simple y logística**

| | *Lineal simple* | *Logística* |
|---|---|---|
| DIFERENCIAS | | |
| Escala de medición de la variable de respuesta: $Y$ | Continua | Dicotómica |
| Distribución probabilística del término aleatorio: $Y\|X$ | Normal: $\mathcal{N}(\mu, \sigma^2)$ | Bernoulli($p$) |
| Parámetros desconocidos | $\mu$, $\sigma^2$ | $P$ |
| Parámetros de interés: media de $Y\|X$ | $E[Y\|X] = \mu$ | $E[Y\|X] = p$: probabilidad de ocurrencia del evento |
| Rango de valores del parámetro de interés | $\mu \in (-\infty, \infty)$ | $P \in [0, 1]$ |
| Modelo propuesto | Recta | POR INVESTIGAR |
| SEMEJANZAS | | |
| Objetivos | • Estimar (cuantificar) la asociación entre una o más variables independientes y una variable de respuesta.<br>• Predicción del valor promedio del parámetro de interés en diversos escenarios. | |



Hasta ahora se han discutido las primeras diferencias señaladas en el cuadro I. Reflexiónese ahora sobre los parámetros de interés en cada modelo. En un modelo de regresión lineal, el parámetro de interés es la media de la variable de respuesta, dados valores fijos de las variables independientes, la cual puede tomar cualquier valor real. En contraste, el parámetro de interés en un modelo logístico es la probabilidad de ocurrencia del evento que, por definición, sólo puede tomar valores entre 0 y 1, inclusive. Mientras que el modelo propuesto en una regresión lineal simple es una recta que toma valores en $(-\infty, \infty)$, coincidente con el rango de valores del parámetro de interés, ahora habrá que pensar en otro modelo que sólo tome valores en $[0, 1]$ igual que el rango de la probabilidad que quiere modelarse. Para esto se requiere de una transformación ingeniosa que permita modelar la asociación entre una variable independiente y la probabilidad de ocurrencia de un evento de una manera continua y acotada en $[0, 1]$: un momio es una expresión que depende del parámetro de interés cuyo rango de posibles valores es el intervalo $[0, \infty)$.

De la función logaritmo natural se sabe que puede aplicarse únicamente a valores en el intervalo $(0, \infty)$ pero de ella se obtiene cualquier número real (igual que una recta). Además, tiene la propiedad de ser una función monótona creciente, es decir, tiene un comportamiento ascendente en todo su recorrido, lo que la hace atractiva como modelo de una relación dosis-respuesta.

¿Qué pasa si se combinan estas dos ideas?, ¿se transforma logarítmicamente un momio y se modela como una función lineal? Piénsese en un caso simple en el que sólo se tenga una variable independiente. Defínase la transformación logit($P$) como

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X \tag{2}$$

Puesto que el parámetro de interés es el valor esperado de la probabilidad de ocurrencia del evento ($P$), es necesario despejarlo por medio de la transformación inversa de la función logaritmo, la exponencial. Después de hacer un poco de álgebra elemental, se llega a lo que se conoce como la **función logística**. Para hacer especial énfasis de que esta función modela la probabilidad de ocurrencia del evento en una situación específica definida por la variable $X$, se denota esta probabilidad como $P(x)$

$$P(\texttt{evento = 1} \mid X = \text{x}) = P(x) = \frac{\exp\left(\beta_0 + \beta_1 x\right)}{1 + \exp\left(\beta_0 + \beta_1 x\right)} = \frac{1}{1 + \exp\left[-\left(\beta_0 + \beta_1 x\right)\right]} \tag{3}$$

Véase qué forma tiene esta función en el caso específico que $\beta_0 = 0$ y $\beta_1 = 1$. Analícese la figura 2, por medio de algunas observaciones:

- La figura 2 representa la asociación entre una exposición continua centrada en 0 y la probabilidad de desarrollar el evento. Cuando esta exposición toma el valor 0, la probabilidad de ocurrencia del evento es igual a ½.
- La función logística es asintótica a 0 y a 1, es decir, se acerca progresivamente a estos valores, pero nunca los alcanza.
- La relación entre la exposición y la probabilidad de ocurrencia del evento es una relación monótona creciente.



En las figuras 3 y 4, se muestra cómo, por medio de los parámetros $\beta_0$ y $\beta_1$, la curva logística puede flexibilizarse para modelar asociaciones con exposiciones de mayor o menor riesgo (figu-

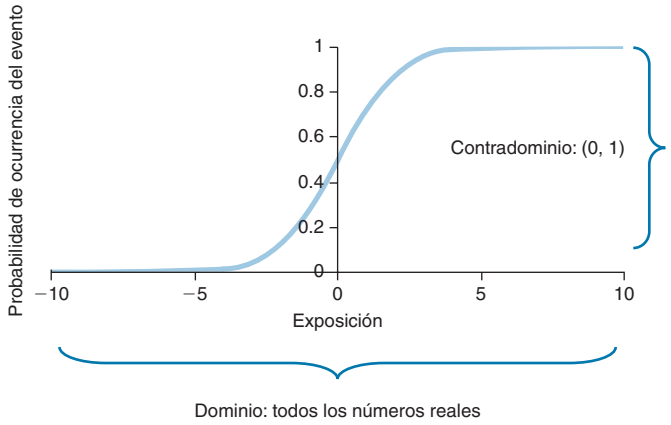



**Figura 2** **Función logística con parámetros β**

ra 3: $\beta_1 > 0$) e, incluso, protectoras (figura 4: $\beta_1 < 0$), es decir, que conforme aumenta la exposición, aumenta o disminuye la probabilidad de ocurrencia del evento, respectivamente.

El valor del coeficiente $\beta_1$ refleja la inclinación de la curva, la cual modela la fuerza de la asociación entre la exposición y la probabilidad de ocurrencia del evento. El coeficiente $\beta_0$ juega un papel equivalente a la ordenada al origen en un modelo de regresión lineal: modelos logísticos con el mismo $\beta_1$, pero diferente $\beta_0$, representan curvas paralelas (misma fuerza de asociación), pero desplazadas en el eje X (figuras 3 y 4, primera y cuarta curvas). Los valores negativos de $\beta_1$ corresponden a situaciones análogas, pero para exposiciones protectoras (figura 4).

Una vez que se ha explorado el comportamiento de la función logística, se pasa a entender cómo se relaciona con la razón de momios. Por facilidad, piénsese en una variable de exposición dicotómica y, posteriormente, hágase una generalización del razonamiento a variables discretas y continuas. En una tabla la probabilidad de ocurrencia ($P$), se resume el evento bajo las cuatro combinaciones posibles y se calculan los momios de desarrollar el evento en cada una de las categorías de la exposición:

| | | Exposición | |
|---|---|---|---|
| | | $X = 1$ | $X = 0$ |
| Evento | $Y = 1$ | $P(1)$ | $P(0)$ |
| | $Y = 0$ | $1 - P(1)$ | $1 - P(0)$ |



$$\text{momio}_{\text{expuesto=1}} = \frac{P(1)}{1-P(1)}$$

$$\text{momio}_{\text{expuesto=0}} = \frac{P(0)}{1-P(0)}$$

$$RM = \frac{P(1)\big/1-P(1)}{P(0)\big/1-P(0)}$$

Por medio del modelo logístico (ecuación 3), $P(1)$ y $P(0)$ se expresarían de la siguiente manera:

$$P(1) = \frac{\exp\left[\beta_0 + \beta_1(1)\right]}{1+\exp\left[\beta_0 + \beta_1(1)\right]} = \frac{\exp\left(\beta_0 + \beta_1\right)}{1+\exp\left(\beta_0 + \beta_1\right)}$$

$$P(0) = \frac{\exp\left(\beta_0\right)}{1+\exp\left(\beta_0\right)}$$

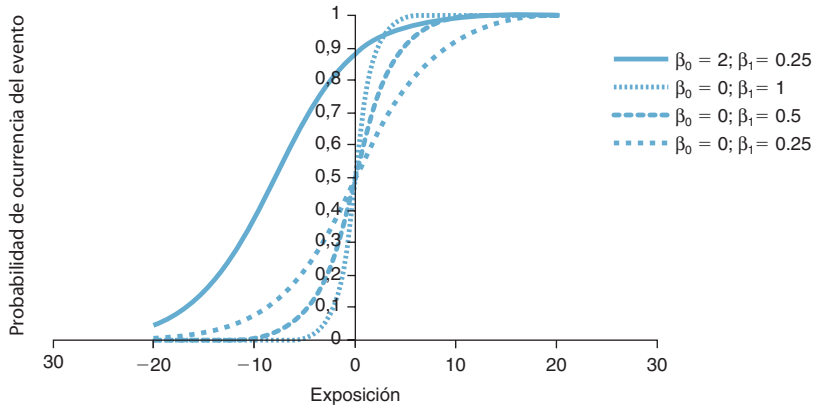



**Figura 3** **Función logística con $\beta_0 \geq 0$ y $\beta_1 > 0$**

de donde puede deducirse con álgebra elemental que

$$1 - P(1) = \frac{1}{1 + \exp\left(\beta_0 + \beta_1\right)}$$

$$1 - P(0) = \frac{1}{1 + \exp\left(\beta_0\right)}$$



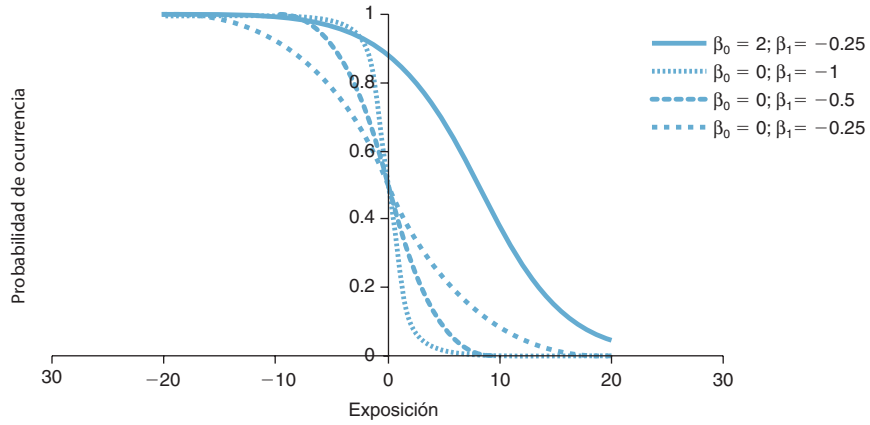



**Figura 4** **Función logística con $\beta_0 \geq 0$ y $\beta_1 < 0$**

y utilizando la fórmula de la *RM* (ecuación 1), se tiene que

$$RM = \frac{P(1)/1-P(1)}{P(0)/1-P(0)} = \frac{\exp(\beta_0+\beta_1)\left[1+\exp(\beta_0+\beta_1)/1+\exp(\beta_0+\beta_1)\right]}{\exp(\beta_0)\left[1+\exp(\beta_0)/1+\exp(\beta_0)\right]} = \exp(\beta_1)$$

entonces

$$\begin{aligned} RM &= \exp(\beta_1) \\ \ln(RM) &= \beta_1 \end{aligned} \tag{4}$$

Esto permite concluir que existe una estrecha relación entre la razón de momios y el coeficiente de regresión que se obtiene por medio de una regresión logística. ¿Qué quiere decir esto en el ejemplo de `osteoporosis` y `thr` anterior y cómo podría calcularse esta *RM* en STATA con el modelo logístico?

```
logistic osteoporosis thr

Logistic regression                                 Number of obs   =        434
                                                    LR chi2(1)      =       4.09
                                                    Prob > chi2     =     0.0431
Log likelihood = -229.78033                         Pseudo R2       =     0.0088
------------------------------------------------------------------------------
osteoporosis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .5508782   .1695377    -1.94   0.053     .3013635    1.006979
------------------------------------------------------------------------------
logit osteoporosis thr, nolog

Logit estimates                                     Number of obs   =        434
                                                    LR chi2(1)      =       4.09
                                                    Prob > chi2     =     0.0431
Log likelihood = -229.78033                         Pseudo R2       =     0.0088
------------------------------------------------------------------------------
osteoporosis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |  -.5962415    .307759    -1.94   0.053    -1.199438     .006955
       _cons |  -1.114549   .1264941    -8.81   0.000    -1.362473    -.866625
------------------------------------------------------------------------------
```

Estas corridas se relacionan con lo visto a través de las siguientes expresiones. Como puede observarse, la *RM* que se obtiene es la misma que se había calculado con la fórmula (1).

$$\hat{P}(\text{osteoporosis} = 1) = \frac{\exp(-1.115 - 0.596 \times thr)}{1 + \exp(-1.115 - 0.596 \times thr)}$$
$$\widehat{RM} = \exp(-0.596) = 0.551 \tag{5}$$

## ESTIMACIÓN PUNTUAL Y POR INTERVALOS

Lo obtenido en la sección anterior, al generar un modelo de regresión logística en la población de estudio, fue un modelo estimado. Se estimaron los parámetros $\beta_0$ y $\beta_1$ por medio de los estimadores muestrales $\hat{\beta}_0 = -1.115$ y $\hat{\beta}_1 = -0.596$, los cuales, para diferenciarlos de los primeros, se acostumbra ponerlos debajo del símbolo ˆ, que se denomina *gorro.* A continuación se verán algunos usos del modelo. Si el diseño de estudio permitiera calcular el *riesgo* (probabilidad) de ocurrencia de un evento, bastaría con sustituir en las variables correspondientes el escenario hipotético en el cual quisiera calcularse dicho riesgo. En un diseño transversal, podría estarse interesado en el cálculo de la prevalencia.

Para estimar la prevalencia de osteoporosis en el grupo de mujeres que no usa `thr` en el ejemplo trabajado, tendría que sustituirse en el modelo (5) `thr`=0:

$$\hat{P}\left[\text{osteoporosis} = 1 \mid thr = 0\right] = \frac{\exp(-1.115)}{1 + \exp(-1.115)} = 0.25$$

Análogamente, la prevalencia estimada para las mujeres que sí acostumbran usar `thr`, podría obtenerse de la misma expresión, sustituyendo `thr`=1:

$$\hat{P}\left[\text{osteoporosis} = 1 \mid thr = 1\right] = \frac{\exp(-1.115 - 0.596)}{1 + \exp(-1.115 - 0.596)} = 0.15$$

Hasta ahora se ha visto que el modelo de regresión logística permitió, no sólo estimar la asociación entre la exposición y la probabilidad de ocurrencia del evento de interés por medio de la *RM*, sino que, además, se estimaron las prevalencias en cada una de las categorías de exposición. No obstante, estos valores corresponden a los que se generaron por medio de una muestra de la población, por lo que se sabe que están sujetos a la variabilidad inherente al proceso de muestreo. Es por ello que, siguiendo los procesos habituales en la estimación estadística, podrían reportarse estos coeficientes estimados con su correspondiente intervalo de confianza.

Como puede verse en las salidas generadas en STATA en la sección anterior, basta fijar la atención en el extremo derecho del renglón correspondiente a la variable `thr` y ahí se encontrará el intervalo de confianza de 95% para la *RM*: (0.301, 1.007). Al ver ahora la forma **logit** del mode-

lo, puede notarse que basta exponenciar los extremos del intervalo de confianza del coeficiente de regresión para obtener el intervalo para la *RM*:

$$\exp(-1.199) = 0.301$$
$$\exp(0.007) = 1.007$$

El cálculo de los intervalos de confianza se realizó bajo el supuesto de que los estimadores de $\beta_0$ y $\beta_1$ siguen una distribución asintóticamente normal:[2]

$$\hat{\beta}_i \overset{a}{\sim} N(\beta_i, Var[\beta_i])\ i = 0, 1$$

por ello, basta aplicar la fórmula (6) para obtener los intervalos con un 95% de confianza:

$$\begin{aligned} \hat{\beta}_0 &\pm 1.96 \times EE(\hat{\beta}_0) \\ \hat{\beta}_1 &\pm 1.96 \times EE(\hat{\beta}_1) \end{aligned} \qquad (6)$$

donde $EE(\hat{\beta}_i)$ se refiere al error estándar $(\sqrt{\widehat{\text{Var}}(\hat{\beta}_i)})$ del estimador correspondiente. Para calcular el intervalo de 95% de confianza para la *RM*, hay que aplicar la transformación exponencial a los extremos del intervalo, es decir,

$$\exp(\hat{\beta}_1 \pm 1.96 \times EE[\hat{\beta}_1])$$

Numéricamente, el intervalo de confianza de 95%, en este ejemplo, se calcularía

$$\exp(-0.596 \pm 1.96 \times 0.308)$$

donde fácilmente puede verse que se llega al intervalo de confianza generado automáticamente por STATA. Cabe señalar que la fórmula (6) es específica para un intervalo de 95%; pero si deseara calcularse este intervalo a cualquier otro nivel de confianza $(1 - \alpha) \times 100\%$, bastaría con reemplazar el 1.96 por el cuantil de una distribución normal estándar $Z_{(1-\alpha/2)}$.

Lo hasta aquí desarrollado se ha centrado en una variable de exposición dicotómica, por medio de una clasificación de los sujetos como expuestos o no expuestos. No obstante, sería deseable tener una medida de la exposición más fina, por ejemplo, en no expuestos, poco expuestos y muy expuestos o, incluso, refinando esta medición hasta hacerla continua.

Si la variable independiente tuviera más de dos categorías, es decir, si fuera politómica, habría que revisar el capítulo sobre regresión lineal y recordar que este problema se aborda por medio de variables indicadoras para cada una de las categorías de exposición. Se incluyen en el modelo tantas variables indicadoras como categorías de exposición, menos una, la cual se utilizará como categoría de referencia. El coeficiente estimado mediante la regresión logística para cada una de las variables indicadoras, tendría la misma interpretación que en la situación de exposición dicotómica, todas ellas con una referencia común: la categoría que decida dejarse fuera del modelo.

Siguiendo con el ejemplo de `osteoporosis` y uso de terapia hormonal de reemplazo (`thr`), supóngase que se cuenta con la variable `tiemtsh_cat`, la cual contiene información sobre el tiempo de uso de `thr` agrupada en las siguientes categorías: 0: no usa `thr`, 1: tiempo de uso ≤3 años; 2: más de tres años. El resumen de esta información se encuentra en el siguiente cuadro:

```
tab tiemtsh_cat

  tiemtsh_cat |      Freq.     Percent        Cum.
--------------+-----------------------------------
No uso de THR |        336       77.42       77.42
      <=3años |         67       15.44       92.86
      >3 años |         31        7.14      100.00
--------------+-----------------------------------
        Total |        434      100.00
```

Como se sabe, los números 0, 1 y 2 sirven meramente como etiquetas ordenadas de tres categorías de la variable `tiemtsh_cat`, pero no reflejan una cualidad numérica, en consecuencia, se trata de una variable ordinal, y para evaluar la asociación entre cada una de estas categorías y la variable resultado `osteoporosis`, hay que generar las correspondientes variables indicadoras de cada categoría:

```
tab tiemtsh_cat, gen(tiemtsh_cat)

tiemtsh_cat |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        336       77.42       77.42
          1 |         67       15.44       92.86
          2 |         31        7.14      100.00
------------+-----------------------------------
      Total |        434      100.00
```

La opción `gen` del comando `tab`, permite la generación automática de las variables indicadoras correspondientes, las cuales pueden visualizarse por medio de comandos básicos descriptivos:

```
desc tiemtsh_cat1 tiemtsh_cat2 tiemtsh_cat3

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
tiemtsh_cat1    byte   %8.0g                  tiemtsh_cat==     0.0000
tiemtsh_cat2    byte   %8.0g                  tiemtsh_cat==     1.0000
tiemtsh_cat3    byte   %8.0g                  tiemtsh_cat==     2.0000
```

```
tab tiemtsh_cat1

tiemtsh_cat |
         == |
     0.0000 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         98       22.58       22.58
          1 |        336       77.42      100.00
------------+-----------------------------------
      Total |        434      100.00

tab tiemtsh_cat2

tiemtsh_cat |
         == |
     1.0000 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        367       84.56       84.56
          1 |         67       15.44      100.00
------------+-----------------------------------
      Total |        434      100.00
```



```
tab tiemtsh_cat3

tiemtsh_cat |
         == |
     2.0000 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        403       92.86       92.86
          1 |         31        7.14      100.00
------------+-----------------------------------
      Total |        434      100.00
```

Una vez que se han construido estas variables indicadoras, es necesario escoger la categoría de referencia y proceder entonces al cálculo de la *RM* correspondiente para cada una de ellas. Desde el punto de vista estadístico, la elección de esta categoría de referencia es irrelevante; no obstante, siguiendo criterios relacionados con interpretación biológica y facilidad para la divulgación de los resultados, se escogerá como referencia aquella categoría de mujeres que no usan `thr`, equivalentemente, `tiemtsh_cat=0` o `tiemtsh_cat1=1`. Por medio de la generación de un modelo de regresión logística, se obtiene:

```
logistic osteoporosis tiemtsh_cat2 tiemtsh_cat3

Logistic regression                                     Number of obs   =        434
                                                        LR chi2(2)      =       4.64
                                                        Prob > chi2     =     0.0980
Log likelihood = -229.50288                             Pseudo R2       =     0.0100
------------------------------------------------------------------------------
osteoporosis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
tiemtsh_cat2 |   .4729954   .1797088    -1.97   0.049     .2246225    .9960028
tiemtsh_cat3 |   .7315663   .3452089    -0.66   0.508     .2901294    1.844657
------------------------------------------------------------------------------
```

Esto indica que la *RM* de tener un diagnóstico positivo para `osteoporosis` en el grupo de mujeres que han estado utilizando `thr` por un máximo de tres años (`tiemtsh_cat2=1`), en relación con las que no lo usan (`tiemtsh_cat1=1`), es 0.47. La correspondiente *RM* para el grupo que ha usado más de 3 años `thr` (`tiemtsh_cat3=1`), en relación con las que no lo usan, es 0.73. Obsérvese que la interpretación de ambas *RM* debe de hacerse utilizando la misma categoría de referencia y que dado que estamos en un estudio transversal, esta *RM* se acostumbra llamar razón de momios de prevalencia.

La pregunta ahora sería, ¿cómo cambia la interpretación de una *RM* (o razón de momios de prevalencia, dependiendo del diseño en cuestión) cuando se tiene una variable de exposición numérica, ya sea continua o discreta? En realidad, hasta ahora se ha interpretado la *RM* que reporta un modelo de regresión logística con exposición dicotómica (utilizando categorías 1 y 0 para el grupo expuesto y no expuesto, respectivamente), como la comparación de momios (hecha por medio de una razón) entre categorías identificadas por códigos numéricos que difieren en una unidad. En otras palabras, se compararon los momios de dos subpoblaciones: aquella identificada como 1, dividida entre el momio de la subpoblación identificada con un 0. Siguiendo esta misma lógica, puede generalizarse rápidamente que la *RM*, asociada a una variable numérica, se interpretaría exactamente de la misma manera. La *RM* indica qué tanto más grande o más pequeño se espera que sea el momio, al comparar dos unidades sucesivas de la variable independiente.

Continuando en el contexto del ejemplo anterior, y en aras de tener una medición más refinada de la variable de exposición, quiere recuperarse la información sobre el número de años que la participante lleva usando `thr`. La búsqueda es exitosa y está contenida en la variable `tiemtsh` (número de años de uso de `thr`).

```
summ tiemtsh

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     tiemtsh |       434    .6920737    1.859769          0         13
```

```
logistic osteoporosis tiemtsh

Logistic regression                                     Number of obs   =        434
                                                        LR chi2(1)      =       1.33
                                                        Prob > chi2     =     0.2483
Log likelihood = -231.15873                             Pseudo R2       =     0.0029
------------------------------------------------------------------------------
osteoporosis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tiemtsh |   .9231384   .0682383    -1.08   0.279     .7986309    1.067057
------------------------------------------------------------------------------
```

Como puede observarse, la media entre las 434 participantes es 0.69 años con un rango de cero a 13 años. Al comparar mujeres que difieren en un año de uso de `thr`, puede verse que aquellas que llevan un año más de uso de `thr` tienen 9% ($1 \div 0.92 = 1.09$) menos posibilidad de tener un diagnóstico positivo de `osteoporosis`. Un supuesto importante que debe resaltarse es que este modelo supone que esta disminución es constante a través de todo el rango de observación de la variable `tiemptsh`, es decir, el resultado de la comparación de los momios es la misma si se comparan subpoblaciones de mujeres que llevan un año usando `thr`, en relación con las que nunca la han usado, que si comparamos aquellas con seis años en relación con las que la han utilizado por cinco años; aquellas con diez años en relación con las de nueve, etcétera.



Ahora, el camino natural a seguir sería evaluar si la relación estimada podría confundirse por la falta de ajuste de otras variables tales como edad, índice de masa corporal, actividad física o cualquier otro confusor posible, es decir, se ha llegado al modelo de regresión logística múltiple.

## MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE

Una vez presentado el modelo de regresión logística con una sola variable independiente, la extensión al modelo múltiple que incluye más predictores es inmediata. Para el caso en que se tengan $k$ variables independientes, la transformación **logit**($P$) y la función logística definidas en las expresiones (2,3) se extienden a la forma:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \qquad (7)$$

$$P(x) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)} = \frac{1}{1 + \exp\left[-\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)\right]} \qquad (8)$$

Como se mencionó en la sección anterior, no existe ninguna restricción sobre la escala de medición de las variables independientes: pueden ser dicotómicas, numéricas discretas o continuas. Para el caso en que se tenga una variable independiente categórica u ordinal, la propuesta sería

generar las variables indicadoras para cada una de las categorías de la variable, por medio de un razonamiento completamente análogo al expuesto en la sección anterior.

Para que los coeficientes de regresión estimados sigan teniendo la misma interpretación que en el caso simple, hay que considerar algunos aspectos que permitan que la deducción de la ecuación (4) siga siendo válida en un contexto donde existen múltiples variables independientes. Específicamente, si quiere evaluarse la asociación entre $X_1$ y la probabilidad de ocurrencia del evento, es importante hacer notar que las variables independientes restantes tienen que permanecer constantes. Aprovéchese esta deducción para introducir también la expresión correspondiente para el caso en que $X_1$ sea cualquier variable numérica, no necesariamente dicotómica:

$$\hat{RM} = \frac{\hat{P}(x_1+1)\,/\,1-\hat{P}(x_1+1)}{\hat{P}(x_1)\,/\,1-\hat{P}(x_1)}$$

Hasta ahora se ha hecho referencia a la función exponencial, como $\exp(x)$; no obstante, dada la longitud de los argumentos que se requieren en las siguientes expresiones, se utilizará su forma alternativa $e^x$.

$$\frac{\dfrac{e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)}}{1+e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)}} \Big/ \dfrac{1}{1+e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)}}}{\dfrac{e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)}}{1+e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)}} \Big/ \dfrac{1}{1+e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)}}} =$$

$$\frac{e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)}\left[1+e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)} \Big/ 1+e^{(\hat{\beta}_0+\hat{\beta}_1\times(x_1+1)+\ldots+\hat{\beta}_k x_k)}\right]}{e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)}\left[1+e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)} \Big/ 1+e^{(\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k)}\right]} = e^{\hat{\beta}_1}$$

Como puede verse, la cancelación de términos es posible porque se partió del supuesto de que las variables $X_2$, $X_3$,…, $X_k$, permanecen sin cambio y sólo se aumentó en una unidad la variable $X_1$, cuya asociación con la variable de respuesta es la que se evaluará. Asimismo, se puede notar que el incremento de una unidad asignado a $X_1$ no demanda la necesidad de que éste se haya dado entre el 0 y el 1, como es el caso de variables dicotómicas, sino que puede ser en cualquiera otra parte del rango de la variable.

Ahora se quiere evaluar la asociación entre el uso `thr` y la probabilidad de tener un diagnóstico positivo de `osteoporosis`, ajustando por dos posibles confusores: edad (`edad`) e índice de masa corporal (`imc`); por lo que se genera el siguiente modelo de regresión logística, el cual podría quedar expresado de manera completamente equivalente en cualquiera de las siguientes dos expresiones que corresponden, respectivamente, a su forma logística y en su forma logit.

$$P(\texttt{osteoporosis}=1)=\frac{\exp\{\beta_0+\beta_1\texttt{thr}+\beta_2\texttt{edad}+\beta_3\texttt{imc}\}}{1+\exp\{\beta_0+\beta_1\texttt{thr}+\beta_2\texttt{edad}+\beta_3\texttt{imc}\}}$$

$$\text{logit}\left[P(\texttt{osteoporosis}=1)\right]=\beta_0+\beta_1\texttt{thr}+\beta_2\texttt{edad}+\beta_3\texttt{imc}$$

El modelo estimado, entonces, se genera con la siguiente salida:

```
logistic osteoporosis thr edad imc
Logistic regression                                 Number of obs   =        434
                                                    LR chi2(3)      =     108.84
                                                    Prob > chi2     =     0.0000
Log likelihood = -177.40359                         Pseudo R2       =     0.2348
------------------------------------------------------------------------------
osteoporosis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .3719543   .1348777    -2.73   0.006     .1827375     .757097
        edad |   1.167025   .0231031     7.80   0.000     1.122611    1.213196
         imc |   .8614773   .0294382    -4.36   0.000     .8056692    .9211512
------------------------------------------------------------------------------
```



```
logit osteoporosis thr edad imc, nolog

Logit estimates                                     Number of obs   =        434
                                                    LR chi2(3)      =     108.84
                                                    Prob > chi2     =     0.0000
Log likelihood = -177.40359                         Pseudo R2       =     0.2348
------------------------------------------------------------------------------
osteoporosis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |  -.9889843   .3626191    -2.73   0.006    -1.699705   -.2782639
        edad |   .1544581   .0197965     7.80   0.000     .1156576    .1932586
         imc |  -.1491066   .0341718    -4.36   0.000    -.2160821   -.0821311
       _cons |   -5.51925   1.297091    -4.26   0.000    -8.061501   -2.976999
------------------------------------------------------------------------------
```

De estas salidas, el modelo estimado puede presentarse en cualquiera de las dos formas alternativas siguientes:

$$\hat{P}(\texttt{osteoporosis}=1)=\frac{\exp\left(-5.52-0.99\texttt{thr}+0.15\texttt{edad}-0.15\texttt{imc}\right)}{1+\exp\left(-5.52-0.99\texttt{thr}+0.15\texttt{edad}-0.15\texttt{imc}\right)}$$

$$\text{logit}\left[\hat{P}(\texttt{osteoporosis}=1)\right]=-5.52-0.99\texttt{thr}+0.15\texttt{edad}-0.15\texttt{imc}$$

En la primera salida, se obtienen directamente las *RM* ajustadas por la presencia de las otras covariables, mientras que en la forma alternativa presentada en la segunda corrida, se presentan los coeficientes de regresión estimados. Como se mencionó anteriormente, las *RM* no son más que el resultado de exponenciar el coeficiente de regresión correspondiente.

## PRUEBA DE HIPÓTESIS RELEVANTE

Seguramente, no sólo quiere estimarse la fuerza de la asociación entre la exposición y el evento de interés, sino además, saber si esta asociación es estadísticamente diferente del valor nulo. Dado que el modelo de regresión logística se ha presentado tanto en su forma logit como en su forma logística, es importante destacar que el valor nulo dependerá de la escala correspondiente. En el modelo logit, la ausencia de asociación está representada por un coeficiente de regresión $\beta_i = 0$, $i = 1, 2,\ldots, k$ y lo equivalente en la forma logística será $\exp(\beta_i) = 1$, $i = 1, 2,\ldots, k$. De aquí que el contraste de hipótesis relevante a realizar sería:

$$H_0: \beta_i = 0 \textit{ vs. } H_a: \beta_i \neq 0 \ (i = 1, 2,\ldots, k)$$

o su equivalente

$$H_0: \exp(\beta_i) = 1 \textit{ vs. } H_a: \exp(\beta_i) \neq 1 \ (i = 1, 2,\ldots, k)$$



Recuérdese que, para poder realizar un contraste de hipótesis, es necesaria una estadística de prueba de la que se conozca su distribución probabilística. En la tercera sección de este capítulo se mencionó que la estimación por intervalos de los estimadores de $\beta_i$, se realizó por medio del supuesto de distribución normal asintótica del correspondiente $\hat{\beta}_i$:

$$\hat{\beta}_i \overset{a}{\sim} N(\beta_i, Var[\beta_i]) \quad i=1, 2,\ldots, k$$

por lo que puede utilizarse la siguiente estadística de prueba,*

$$Z = \frac{\hat{\beta}_i}{EE(\hat{\beta}_i)} \overset{a}{\sim} N(0,1)$$

Puesto que esta estadística sigue una distribución normal estándar, la región crítica o de rechazo de la hipótesis nula se construye de manera habitual.

En la práctica, esta prueba de hipótesis está contenida dentro de la salida que genera el paquete estadístico en la columna con el título `P>|z|`, el cual contiene el valor *p* asociado con la prueba en cuestión. Si el valor reportado es menor que el nivel de significancia al que se quiere realizar el contraste ($\alpha$), entonces se rechaza la hipótesis de asociación nula en los parámetros y

* Algunos autores le llaman la estadística de Wald; [2,3] sin embargo, otros le dan este nombre a la estadística que resulta de elevarla al cuadrado, teniendo como distribución resultante una $\chi^2_{(1)}$. [4]

se concluye que, a un nivel de significancia $\alpha$, la asociación correspondiente es "estadísticamente significativa". En general, aunque no necesariamente, se acostumbra realizar este contraste de hipótesis a un nivel de significancia de 0.05, lo cual se traduce en que los valores $p < 0.05$ se interpretan como asociaciones significativas.

El ejemplo que se ha manejado ha sido:

```
logistic osteoporosis thr edad imc
Logistic regression                                 Number of obs   =        434
                                                    LR chi2(3)      =     108.84
                                                    Prob > chi2     =     0.0000
Log likelihood = -177.40359                         Pseudo R2       =     0.2348
------------------------------------------------------------------------------
osteoporosis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |   .3719543   .1348777    -2.73   0.006     .1827375     .757097
        edad |   1.167025   .0231031     7.80   0.000     1.122611    1.213196
         imc |   .8614773   .0294382    -4.36   0.000     .8056692    .9211512
------------------------------------------------------------------------------

logit osteoporosis thr edad imc, nolog

Logit estimates                                     Number of obs   =        434
                                                    LR chi2(3)      =     108.84
                                                    Prob > chi2     =     0.0000
Log likelihood = -177.40359                         Pseudo R2       =     0.2348
------------------------------------------------------------------------------
osteoporosis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         thr |  -.9889843   .3626191    -2.73   0.006    -1.699705   -.2782639
        edad |   .1544581   .0197965     7.80   0.000     .1156576    .1932586
         imc |  -.1491066   .0341718    -4.36   0.000    -.2160821   -.0821311
       _cons |   -5.51925   1.297091    -4.26   0.000    -8.061501   -2.976999
------------------------------------------------------------------------------
```



Como se comentó en las secciones anteriores, de estas salidas puede obtenerse el valor puntual estimado de la *RM* para la variable terapia hormonal de reemplazo (`thr`) ajustado por edad e índice de masa corporal: 0.37 con un intervalo de confianza de 95% de (0.18, 0.76). Ahora se realizará la estadística de prueba para entender cómo se realiza la prueba de hipótesis sobre la significancia de esta asociación. La información necesaria para este cálculo está contenida en la salida bajo las columnas "`Coef`", "`Std. Err.`" y "`z`" de la salida del modelo en su forma logit:

$$Z = \frac{\hat{\beta}_i}{E\hat{E}(\hat{\beta}_i)} = \frac{-0.9890}{0.3626} = -2.73$$

Como el valor calculado de la estadística es $-2.73$ y $|-2.73| = 2.73 > 1.96$, puede concluirse que esta asociación se considera estadísticamente significativa a un nivel de significancia de $\alpha = 0.05$. Más aún, basados en el valor $p$ reportado de 0.006, puede decirse que la probabilidad de que se concluya erróneamente que esta asociación es diferente de la nula es de 0.006 (muy pequeña). La evidencia muestral apoya la asociación protectora de la `thr` sobre el diagnóstico de `osteoporosis`.

En realidad, no es necesario hacer este procedimiento cada vez que desee realizarse esta prueba de hipótesis. Basta con ver el valor $p$ e interpretarlo adecuadamente.

## DIAGNÓSTICO

En esta sección se presentará un breve resumen de las técnicas que se utilizan con mayor frecuencia para el diagnóstico de un modelo de regresión logística, partiendo del supuesto de que se tiene ya un modelo estimado, el cual parece razonablemente satisfactorio. Ahora quiere explorarse qué tan bien describe el modelo la relación entre la variable de respuesta y las variables independientes.

### Prueba del cociente de verosimilitudes

Se utilizarán las estadísticas que se incluyen en la salida de STATA en la parte superior derecha, que se destacan con un recuadro punteado de la corrida anterior. Después de reportar el número de observaciones incluidas en el modelo, aparece la leyenda `LR chi2(3)=108.84` que es el valor de la estadística de prueba que compara la función de verosimilitud*,[5] del modelo propuesto con la del modelo nulo, es decir, aquel que no incluye variables independientes.[6] Estrictamente hablando, la prueba del cociente de verosimilitudes compara la verosimilitud del modelo propuesto con la del modelo saturado, que es aquel que contiene tantos parámetros como número de observaciones haya en la muestra. Sin embargo, lo que hace internamente STATA es comparar tanto el modelo propuesto como el modelo nulo con el modelo saturado, lo que da como resultado que la verosimilitud del modelo saturado se cancele y sea equivalente a comparar directamente las verosimilitudes del modelo propuesto contra la del modelo nulo. Así, la prueba de hipótesis subyacente a la estadística reportada por STATA es:

$$H_0: L_P = L_0 \qquad vs. \qquad H_1: L_P = L_0$$

donde $L_P$ y $L_0$ denotan las funciones de verosimilitud del modelo propuesto y el modelo nulo, respectivamente. La estadística de prueba utiliza el logaritmo de estas funciones y se puede probar que si $H_0$ es cierta, sigue una distribución $\chi^2_{(k)}$:

$$X^2 = 2(\ln L_P - \ln L_0) \sim \chi^2_{(k)}$$

* Se recomienda que el lector revise el concepto de función de verosimilitud en la bibliografía propuesta. Sin embargo, es necesario resaltar que es una función que se busca maximizar en el proceso de estimación; se acostumbra transformarla logarítmicamente dando como resultado lo que se conoce como log-verosimilitud; una vez transformada, toma valores negativos, y los valores "grandes" tanto de la verosimilitud como de la log-verosimilitud pueden considerarse evidencia de un mejor modelo.

Con fines de comprensión, véase de dónde se obtiene este valor en el ejemplo que se ha venido trabajando a lo largo del capítulo. Al correr el modelo nulo, se ve en la parte superior izquierda de la corrida que la función log-verosimilitud toma el valor de −231.82517; por su parte, el valor que alcanza esta función en el modelo propuesto es −177.40359 (nótese que $\ln L_0 = -231.8217 < -177.40359 = \ln L_p$)

```
logit osteoporosis , nolog
Logit estimates                                   Number of obs   =        434
                                                  LR chi2(0)      =      -0.00
                                                  Prob > chi2     =          .
Log likelihood = -231.82517                       Pseudo R2       =    -0.0000
------------------------------------------------------------------------------
osteoporosis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -1.232144   .1148054   -10.73   0.000    -1.457158   -1.007129
------------------------------------------------------------------------------
```

La estadística de prueba se calcularía como

$$=2[-177.40359 - (-231.82517)] = 108.84$$



Puesto que el modelo propuesto incluye tres variables independientes, este valor debería compararse contra una distribución $\chi^2_{(3)}$, que es la diferencia de parámetros a estimar entre los modelos que se están contrastando: el nulo y el propuesto. El renglón siguiente reporta el valor *p* correspondiente a esta prueba que es de una cola (ya que la inclusión de una variable siempre repercutirá en un valor de la función de máxima verosimilitud mayor o igual al modelo sin ella). En nuestro ejemplo, este valor es altamente significativo, lo cual indica que la aportación conjunta de las tres variables a la explicación de la variable de respuesta es estadísticamente significativa. En general, no hay que hacer toda esta deducción, pues basta con interpretar el valor *p* de una manera análoga a como se interpreta en la prueba *F* para la evaluación global de un modelo de regresión lineal presentada en el capítulo correspondiente.

La proporción de incremento de la función log-verosimilitud del modelo propuesto en relación con el modelo nulo se define como pseudo $R^2$. En STATA se reporta en el renglón siguiente a la prueba del cociente de verosimilitudes y se calcula de la siguiente manera:

$$\text{Pseudo } R^2 = \frac{\ln(L_0) - \ln(L_p)}{\ln(L_0)}$$

lo que en el ejemplo se calcularía como:

$$\frac{-231.82517 - (-177.40359)}{-231.82517} = 0.2348$$

Esta estadística tiene varios inconvenientes. Su interpretación no es intuitiva. A diferencia de la $R^2$ del modelo de regresión lineal, ésta se interpreta en términos de la función de verosimili-

tud que no es fácil de expresarse en el lenguaje del problema de estudio. Además, su valor máximo es menor a 1 cuando $J < n$ ($J$ es el número total de patrones de covariables observados en la muestra que se presenta más adelante); la pseudo $R^2$ tiende a reportar valores más bajos que las $R^2$ que se obtienen de modelos lineales, por lo que puede ser desconcertante cuando se presenta ante audiencias más familiarizadas con la regresión lineal. Por lo anterior, no se recomienda el uso de esta estadística, excepto cuando se utilice como un criterio más en la selección de modelos alternativos. La razón por la que se menciona en este capítulo atiende a que el paquete estadístico con el que se trabaja la reporta en su salida.

Adicionalmente a estas estadísticas (cociente de verosimilitudes y pseudo $R^2$), existen otras herramientas para evaluar al modelo, las cuales se basan, esencialmente, en la comparación entre los valores observados y los valores esperados. A diferencia del modelo lineal, es necesario explicar qué significa esta diferencia, ya que el valor observado de la variable dependiente en un modelo de regresión logística es 0 o 1, mientras que el valor esperado es una probabilidad estimada, es decir, un valor que se encuentra en el intervalo (0, 1).*

La idea central de las pruebas que se exponen radica en comparar el número de eventos que se presentaron en un grupo definido de cierta manera, con el número esperado en ese grupo, de acuerdo con el modelo propuesto. En realidad, ésta se presenta ante una variable de tipo binomial en donde el número de "ensayos" (siguiendo el lenguaje que se utiliza comúnmente para definir este tipo de variables aleatorias) sería el número de observaciones en cada uno de estos grupos, y los comúnmente denominados "éxitos" serían las observaciones que desarrollaron el evento (`osteoporosis=1`) en cada uno de ellos. Para el cálculo de los valores esperados en una variable aleatoria binomial, se requiere de una probabilidad de ocurrencia del evento. Así, en este contexto, se propone utilizar la probabilidad estimada por el modelo en cada uno de estos grupos, de tal manera que si el modelo fuera correcto, la diferencia entre los valores observados y los valores esperados debería de ser pequeña. La forma como se definen estos grupos será lo que caracterice las pruebas que se presentan. En este capítulo se revisarán dos pruebas de uso frecuente que pueden obtenerse en el paquete STATA: la prueba *ji* cuadrada de Pearson y la prueba de Hosmer-Lemeshow.

Para poder comprenderlas, es necesario introducir un concepto nuevo y un poco de notación adicional que jugará un papel relevante en su desarrollo. Se utilizará el concepto "patrón de covariables" para describir una combinación única de valores observados de las covariables incluidas en el modelo, esto es, los individuos que tengan los mismos valores observados en todas sus covariables. El número total de patrones de covariables observados en la muestra es $J$; cuando todas las variables independientes del modelo son dicotómicas, el número de patrones de covariables será generalmente mucho menor que el número de observaciones ($J < n$), ya que varios sujetos tendrán la misma combinación de valores observados de las variables independientes. Sin embargo, cuando se incluyen variables independientes numéricas y en particular, continuas, este número podría aproximarse mucho al tamaño de muestra. Véase qué ocurre en el ejemplo que se ha trabajado.

El modelo propuesto incluye tres covariables: `thr`, `edad` e `imc`. Obsérvese la información correspondiente de algunas de las mujeres participantes en el estudio.

* Cabe señalar que, aunque el rango posible de valores de una probabilidad es el intervalo [0, 1], la probabilidad estimada con un modelo logístico estará contenida en (0, 1).

```
list folio osteoporosis thr edad imc in 1/10
```

```
     +-----------------------------------------+
     | folio   osteop~s   thr   edad       imc |
     |-----------------------------------------|
  1. |  8564          1     0     41  36.45692 |
  2. |   950          0     0     42  26.50212 |
  3. |  3740          0     0     42  23.49524 |
  4. |  5076          0     0     42  23.83301 |
  5. |  6149          0     1     42  26.02264 |
     |-----------------------------------------|
  6. |  5422          0     0     43  23.55734 |
  7. |  1508          0     1     43  28.04014 |
  8. |   930          0     0     44  35.02926 |
  9. |  3059          0     0     44  23.60128 |
 10. |  9057          1     0     44   30.6302 |
     |                                         |
     +-----------------------------------------+
```

patrón de covariables en ausencia (presencia) de imc en el modelo

| | |
|---|---|
| 1 | (1) |
| | (2) |
| 2 | (3) |
| | (4) |
| 3 | (5) |
| | (6) |
| 4 | (7) |
| 5 | (8) |
| | (9) |
| 6 | (10) |

Aquí puede verse que en la información listada se detectan diez patrones de covariables (ver numeración entre paréntesis en la columna derecha de la salida), tantos como observaciones listadas. Esto se debe a que la variable `imc` es continua y no se repite en ninguno de los perfiles definidos por las variables `thr` y `edad`. Si `imc` no hubiera estado incluida en el modelo, el listado anterior habría incluido seis patrones de covariables (ver numeración sin paréntesis en la columna derecha de la salida): el de `edad` =41 y `thr`=0 con una participante (folio 8564), el de `edad`=42 y `thr`=0 con tres participantes (folios 950, 3740 y 5076); el de `edad`=42 y `thr`=1 (folio 6149) y los tres restantes: `edad`=43 con `thr`=0 y `thr`=1 y `edad`=44 con `thr`=0 con uno, uno y tres participantes, respectivamente.



El número de sujetos que comparten el $j$-ésimo patrón de covariables se denotará por $m_j$ ($j = 1, \ldots, J$) y su suma deberá de ser igual al tamaño de la muestra, es decir, $\sum_{j=1}^{J} m_j = n$. En la tabla listada, y siguiendo con el supuesto de que `imc` no hubiera estado en el modelo, $m_1 = 1$, $m_2 = 3$, $m_3$=1, $m_4$=1, $m_5$=1 y $m_6$=3. Dentro de estos $m_j$, se denotará por $y_j$ el número de sujetos en el $j$-ésimo patrón de covariables que presentaron el evento de interés ($y_1 = 1, y_2 = 0, y_3 = 0, y_4 = 0, y_5 = 0, y_6 = 1$), es decir, todos aquellos cuyo valor observado de la variable de respuesta fue igual a 1, y que en este caso significa que tuvieron un diagnóstico positivo de osteoporosis.

### *Ji* cuadrada de Pearson

Para un patrón de covariables específico, se define el residuo de Pearson como la diferencia estandarizada entre el número observado de eventos ($y_j$) y el número esperado correspondiente. Puesto que este patrón de covariables es de tamaño $m_j$ y el modelo estima una probabilidad de ocurrencia del evento de $\hat{p}_j$, tenemos una distribución binomial con parámetros $m_j$ y $\hat{p}_j$; por lo tanto, el valor esperado es igual $m_j\hat{p}_j$ y la varianza igual a $m_j\hat{p}_j(1 - \hat{p}_j)$. De ahí que el residuo de Pearson sea:

$$r\left(y_j, \hat{p}_j\right) = \frac{y_j - m_j \hat{p}_j}{\sqrt{m_j \hat{p}_j (1 - \hat{p}_j)}}$$

La estadística de prueba denominada *ji* cuadrada de Pearson es la suma de los cuadrados de estos residuos, y se utiliza para probar la hipótesis nula de que el modelo ajusta bien a los datos. Bajo el supuesto de que el modelo es correcto en todos los aspectos, esta estadística se distribuye como una *ji* cuadrada con $J - (k + 1)$ grados de libertad (recuérdese que $k$ es el número de variables independientes en el modelo)

$$X^2 = \sum_{j=1}^{J} r^2\left(y_j, \hat{p}_j\right) \sim \chi^2_{j-(k+1)}$$

Sin embargo, esta estadística tiene problemas distribucionales cuando el número de individuos dentro de los patrones de covariables es pequeño (como regla de dedo, se utiliza como criterio un mínimo de cinco observaciones) y $J$ se aproxima mucho al tamaño de la muestra ($J \rightarrow n$), por lo que los valores $p$ calculados en esta situación podrían ser incorrectos y deben interpretarse con cautela. Para estas situaciones, se recomienda la prueba de Hosmer-Lemeshow, que se presenta a continuación.

**Prueba de Hosmer-Lemeshow**

La idea de esta prueba es comparar los valores observados contra los esperados en grupos definidos por las probabilidades estimadas por el modelo. Estos grupos pueden generarse siguiendo dos estrategias: 1) utilizar los percentiles de la probabilidad estimada, y 2) dividir el rango de las probabilidades estimadas por números arbitrarios predefinidos. Con la primera propuesta, los grupos quedarán balanceados en tamaño, mientras que con la segunda estrategia esto podría no suceder.

Una vez que se han definido estos $g$ grupos, se genera una tabla de $2 \times g$, en donde un renglón será el correspondiente para los sujetos que presentaron el evento ($Y = 1$) y el otro corresponderá a aquellos que no lo presentaron ($Y = 0$). Para cada celda, se tendrá un valor observado al que se le llamará $o_k$, correspondiente a la frecuencia de eventos del renglón en el grupo respectivo. El valor esperado en cada una de las celdas se calculará multiplicando el número de sujetos en ese grupo ($n'_k$) por la probabilidad promedio en los $c_k$ diferentes patrones de covariables que conforman el $k$-ésimo grupo; para el renglón donde $Y = 1$, se suman las probabilidades estimadas de todos los sujetos en el grupo, mientras que para el renglón $Y = 0$, se sumará el complemento a uno de las probabilidades estimadas. Finalmente, la estadística de prueba de Hosmer-Lemeshow se obtiene haciendo una prueba *ji* cuadrada de Pearson sobre la tabla $2 \times$ g:

$$\hat{C} = \sum_{k=1}^{g} \frac{\left(o_k - n'_k \bar{p}_k\right)^2}{n'_k \bar{p}_k \left(1 - n'_k \bar{p}_k\right)}$$

donde

$$\bar{p}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{p}_j}{n'_k}$$

En 1980,[7] Hosmer y Lemeshow probaron mediante simulaciones que, cuando $J = n$ y el modelo es correcto, la distribución de esta estadística de prueba se aproxima muy bien a una *ji* cuadrada con $g - 2$ grados de libertad. Posteriormente, en 1988,[8] los mismos autores probaron que la estrategia de agrupamiento definida por los percentiles de las probabilidades estimadas era preferible a la de puntos de corte arbitrarios en el sentido de su aproximación a la distribución $\chi^2_{(g-2)}$.

El comando **`lfit`** de STATA cuenta el número de patrones de covariables, calcula la estadística *ji* cuadrada de Pearson y proporciona el valor *p* de la prueba correspondiente. En el ejemplo, se ve que el número de patrones de covariables ($J = 431$) se aproxima mucho al tamaño de la muestra ($n = 434$), por lo que el valor *p* de 0.0041, que apoya el rechazo de la hipótesis nula de un buen ajuste del modelo, debe de tomarse con reserva.

```
lfit

Logistic model for osteoporosis, goodness-of-fit test

      number of observations =       434
 number of covariate patterns =       431
           Pearson chi2(427) =       508.21
                 Prob > chi2 =         0.0041 ←
```



Si se agrega la opción **`group(#)`**, STATA calculará la estadística de prueba Hosmer-Lemeshow, utilizando los percentiles de la probabilidad estimada para generar el número de grupos solicitados; con la opción **`table`** se desplegará, adicionalmente, el cuadro generado para realizar la prueba. Siguiendo la recomendación de estos autores, se generarán diez grupos. Aquí puede verse el cuadro con la información generada para cada grupo: el número de eventos observados y esperados, así como el número de sujetos que no desarrollaron el evento y el valor esperado para ellos estimado por el modelo. La última columna del cuadro muestra el número total de observaciones en cada uno de los grupos. Por último, se ve que el valor *p* de 0.5715 apoya la hipótesis nula y sugiere que el modelo propuesto se ajusta razonablemente bien a los datos.

```
lfit, group(10) table

Logistic model for osteoporosis, goodness-of-fit test
  (Table collapsed on quantiles of estimated probabilities)
  +--------------------------------------------------------+
  | Group |   Prob | Obs_1 | Exp_1 | Obs_0 | Exp_0 | Total |
  |-------+--------+-------+-------+-------+-------+-------|
  |     1 | 0.0323 |     2 |   0.9 |    42 |  43.1 |    44 |
  |     2 | 0.0650 |     2 |   2.2 |    41 |  40.8 |    43 |
  |     3 | 0.0899 |     4 |   3.3 |    40 |  40.7 |    44 |
  |     4 | 0.1179 |     3 |   4.4 |    40 |  38.6 |    43 |
  |     5 | 0.1506 |     2 |   5.8 |    41 |  37.2 |    43 |
```

```
|-------+--------+-------+-------+-------+-------+-------|
|     6 | 0.1969 |    10 |   7.8 |    34 |  36.2 |    44 |
|     7 | 0.2554 |     8 |   9.7 |    35 |  33.3 |    43 |
|     8 | 0.3414 |    15 |  13.0 |    29 |  31.0 |    44 |
|     9 | 0.5295 |    19 |  18.9 |    24 |  24.1 |    43 |
|    10 | 0.9756 |    33 |  32.0 |    10 |  11.0 |    43 |
+-------------------------------------------------------+

      number of observations =       434
            number of groups =        10
    Hosmer-Lemeshow chi2(8) =      6.68
                Prob > chi2 =      0.5715   <----
```

Hasta ahora se ha presentado el modelo de regresión logística, sus posibles usos e interpretaciones, y un breve resumen de técnicas diagnósticas. Sin embargo, se recomienda que, al generarse un modelo de este tipo, se revisen en forma exhaustiva los criterios necesarios para generar un modelo, se realice un diagnóstico profundo del mismo que incluya el estudio individual de los residuos, las medidas de influencia y la generación de intervalos de predicción, entre otros.

A manera de conclusión, se presenta una sección en la que se vincula esta metodología de modelaje estadístico con el diseño epidemiológico que generó la información.

## REFLEXIONES SOBRE EL USO DE LA REGRESIÓN LOGÍSTICA EN RELACIÓN CON EL DISEÑO DE ESTUDIO



¿Qué interpretación tiene el coeficiente $\beta_0$ en un modelo de regresión logística con una variable de exposición dicotómica $X$ (0: no expuesto, 1: expuesto)?

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

Cuando $X$ es igual a cero, $\beta_0$ es el logaritmo natural del momio del evento:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1(0) = \beta_0$$

por lo tanto exp $(\beta_0)$ representa el momio del evento cuando $X = 0$. Nótese que no es una *RM*. Si $X$ es una variable de exposición dicotómica, exp $(\beta_0)$ representa el momio de la enfermedad en el grupo no expuesto.

En un estudio de casos y controles, la selección de la muestra se basa en la presencia o ausencia del evento de interés; esta muestra no es representativa de la población objetivo, ya que la selección de casos busca la representatividad de la totalidad de los casos que ocurrieron en la población fuente en un tiempo determinado, y la selección de controles busca la representatividad de la población fuente que generó los casos. Esta estrategia de muestreo con base en el evento

conduce a que los momios muestrales difirieran significativamente de los momios de la población fuente. Los momios del evento en las categorías de la exposición se fijan de manera arbitraria, de acuerdo con el número de controles por caso que se hayan seleccionado, por lo tanto, exp ($\beta_0$) no tiene validez. Esto último se dice en el sentido más estricto del concepto de validez epidemiológica, ya que no es representativo de lo que ocurre en la población objetivo. En un muestreo de este tipo, se busca que los momios de la exposición en los grupos de casos y de controles seleccionados representen adecuadamente a los momios respectivos de la población fuente.

***Ejemplo:***
Supóngase que se estudia una cohorte de 5 050 individuos, divididos en dos categorías de exposición: hipertensos (expuestos) y normotensos (no expuestos). El evento de interés: la ocurrencia de infarto agudo del miocardio (IAM). La realidad de lo que sucedió en ambas categorías de la exposición fue la siguiente:

| | | **Exposición: hipertensión** | | |
|---|---|---|---|---|
| | **Diagnóstico** | 1: Sí | 0: No | **Total** |
| **Evento: IAM** | 1 : dx positivo | 30 | 20 | 50 |
| | 0 : dx negativo | 2 000 | 3 000 | 5 000 |
| | Total | 2 030 | 3 020 | 5 050 |



momio de ser caso en los expuestos: 30/2,000 = 0.015
momio de ser caso en los no expuestos: 20/3,000 = 0.0066
momio de la exposición en los casos: 30/20 = 1.5
momio de la exposición en los no casos: 2 000/3 000 = 0.66

$$RM = \frac{\text{momio}_{\text{expuesto=1}}}{\text{momio}_{\text{expuesto=0}}} = \frac{\frac{30}{20}}{\frac{20}{30}} = 2.25$$

Como se cuenta con recursos limitados para entrevistar a toda la cohorte, se decidió realizar un estudio de casos y controles, contrastando la totalidad de los casos ($n = 50$) con una muestra aleatoria de la población (controles), con una fracción de muestreo de 0.01 en donde por simplicidad, no se consideran errores de muestreo (los controles reproducen las proporciones de la exposición exactamente)

| | | **Exposición: hipertensión** | | |
|---|---|---|---|---|
| | **Diagnóstico** | 1: Sí | 0: No | **Total** |
| **Evento: IAM** | 1 : dx positivo | 30 | 20 | 50 |
| | 0 : dx negativo | 20 | 30 | 50 |
| | Total | 50 | 50 | 100 |

momio de ser caso en los expuestos: 30/20 = 1.5
momio de ser caso en los no expuestos: 20/30 = 0.66
momio de la exposición en los casos: 30/20 = 1.5
momio de la exposición en los controles: 20/30 = 0.66

$$RM = \frac{\text{momio}_{\text{expuesto=1}}}{\text{momio}_{\text{expuesto=0}}} = \frac{\frac{30}{20}}{\frac{20}{30}} = 2.25$$

La particularidad de que la razón de momios sea simétrica hace que las *RM* estimadas en un estudio de casos y controles sean insesgados respecto a los *RM* en la población fuente. Como se observa en el ejemplo: los momios de la exposición en casos y controles son los mismos que los de la población fuente. Por el contrario, los momios del evento en las categorías de la exposición son muy diferentes a los de la cohorte completa (esta diferencia depende multiplicativamente de la fracción de muestreo que se utiliza para seleccionar la muestra de controles). Puesto que el parámetro $\beta_0$ es el momio del evento en los no expuestos, se concluye que $\beta_0$ no puede estimarse en un estudio de casos y controles. En consecuencia, el riesgo (probabilidad) del evento no puede estimarse para ninguna categoría de la exposición, dado que para obtenerlo es necesario incluir $\beta_0$. Por lo tanto, el parámetro *P* no es estimable en un estudio de casos y controles (a menos que se conozca la fracción de muestreo de los controles).

En cambio, $\beta_0$ sí puede estimarse adecuadamente en los estudios transversales en los que se utiliza un muestreo aleatorio simple. En estos casos, las celdas están igualmente representadas en la muestra final y las relaciones entre las celdas son iguales a las de la población fuente. Por lo tanto, en estudios transversales, la prevalencia del evento (estadísticamente hablando, la probabilidad de ocurrencia del mismo) sí es estimable.



***Ejemplo:***
Supóngase que se lleva a cabo un estudio transversal que, entre otras cosas, recopiló información sobre el sexo y la presencia de diabetes mellitus (DM). Los datos de toda la población son:

| | | Sexo | | |
|---|---|---|---|---|
| | **Diagnóstico** | Mujeres | Hombres | **Total** |
| **Evento: DM** | 1 : dx positivo | 3 508 | 2 020 | 5 528 |
| | 0 : dx negativo | 8 345 | 12 126 | 14 146 |
| | Total | 11 853 | 14 146 | 25 999 |

momio de ser diabético en el grupo de mujeres: 3 508/8 345 = 0.42
momio de ser diabético en el grupo de hombres: 2 020/12 126 = 0.17
momio de ser mujer en los diabéticos: 3 508/2 020 = 1.74
momio de ser mujer entre los no diabéticos: 8 345/12 126 = 0.69

$$RM = \frac{\text{momio}_{\text{mujeres}}}{\text{momio}_{\text{hombres}}} = \frac{\frac{3\,508}{8\,345}}{\frac{2\,020}{12\,126}} = 2.52$$

Como no es posible hacer un censo de la comunidad, se obtiene una muestra aleatoria sobre la cual se hará el estudio. La fracción de muestreo es de 0.01 y se obtienen las siguientes estimaciones:

| | | Sexo | | |
|---|---|---|---|---|
| | **Diagnóstico** | Mujeres | Hombres | **Total** |
| **Evento: DM** | 1 : dx positivo | 35 | 20 | 50 |
| | 0 : dx negativo | 83 | 121 | 204 |
| | Total | 118 | 141 | 254 |

momio de ser diabético dado que se es mujer: 35/83 = 0.42
momio de ser diabético dado que se es hombre: 20/121 = 0.17
momio de ser mujer dado que se es diabético: 35/20 = 1.75
momio de ser mujer dado que no se es diabético: 83/121 = 0.69



$$\hat{RM} = \frac{\hat{\text{momio}}_{\text{mujeres}}}{\hat{\text{momio}}_{\text{hombres}}} = \frac{\frac{35}{83}}{\frac{20}{121}} = 2.55$$

Como puede observarse, los momios estimados son iguales a los verdaderos. La diferencia de tres centésimas se debe a los errores de redondeo al seleccionar la muestra aleatoria con fracción de muestreo de 0.01. Si con estos datos se propusiera un modelo de regresión logística, donde DM fuera la variable de respuesta y el sexo la variable independiente (sexo = 1 para las mujeres y sexo = 0 para los hombres), exp ($\hat{\beta}_0$) sería la estimación del momio de ser diabético en los hombres. Esta estimación, si no hay errores de muestreo, es igual al parámetro de interés, por lo que es perfectamente válida. Dado que $\beta_0$ sí puede estimarse en forma confiable en este tipo de estudios, es posible calcular la probabilidad (prevalencia) del evento para cualquier categoría de la variable independiente o de exposición.

Finalmente, en los estudios de cohorte, $\beta_0$ también es estimable, pues en estos estudios generalmente sólo pretenden extrapolarse los resultados a la cohorte misma. En consecuencia, el riesgo (probabilidad del evento) es estimable para todas las categorías de la exposición.

## REFERENCIAS


1. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Applied regression analysis and other multivariable methods. 3a. ed. Nueva York: Duxbury Press, 1998.
2. Kleinbaum DG. Logistic regression: a self learning text. 2a. ed. Washington: Springer-Verlag, 2002.
3. Hosmer DW. Lemeshow S. Applied logistic regression. 2a. ed. Washington: John Wiley & Sons, 2000.
4. Collet D. Modelling binary data. Nueva York: Chapman & Hall, 1991.
5. Arnold SF. Mathematical statistics. Englewood Cliffs: Prentice Hall, 1990.
6. Dobson, A.J. An introduction to generalized linear models. Nueva York: Chapman & Hall, 1990.
7. Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. Communications in statistics 1980;A10:1043-1069.
8. Hosmer DW, Lemeshow S, Klar J. A goodness-of-fit testing for multiple logistic regression analysis when the estimated probabilities are small. Biometrical Journal 1988;30:911-924.

# XV Análisis de supervivencia

*Salvador Zamora Muñoz*
*Ma. de Lourdes Flores Luna*

## INTRODUCCIÓN

Los estudios de cohorte son comunes en la investigación epidemiológica. Como se vio con anterioridad, estos estudios se caracterizan porque a los individuos se les sigue a lo largo de un determinado periodo. De acuerdo con la naturaleza de estos estudios, existen diversas técnicas estadísticas para analizar la información recabada. Algunas de ellas se agrupan en el área estadística denominada análisis de supervivencia. El interés en este tipo de análisis radica en estudiar el tiempo que transcurre entre un evento inicial (que determina la inclusión en el estudio de un individuo) y un evento final (genéricamente denominado falla), que define el término del estudio para cada individuo. Al tiempo transcurrido se le denomina tiempo de falla o de supervivencia. En los estudios epidemiológicos, los tipos de eventos finales son, por ejemplo, la muerte, la ocurrencia de alguna enfermedad, la recuperación de algún paciente enfermo, etc. Aunque los estudios de supervivencia son característicos de la investigación epidemiológica, existen otras áreas que presentan fenómenos que pueden analizarse por medio de los métodos del análisis de supervivencia, por ejemplo: en la industria, el tiempo de falla de un componente de alguna máquina o aparato electrónico; en economía, el tiempo de desempleo de una persona económicamente activa; en demografía, el tiempo de duración del matrimonio; en psicología, el tiempo para realizar alguna tarea en una evaluación psicológica, entre otras.

En resumen, el objetivo del análisis de supervivencia es estudiar el tiempo (tiempo de falla o de supervivencia) que transcurre entre los eventos que determinan el inicio y el fin del estudio. Estos métodos comprenden las etapas usuales del análisis estadístico: análisis descriptivo, comparación del tiempo de supervivencia entre distintas poblaciones, evaluación del impacto de algunas covariables en este tiempo de supervivencia por medio de modelos de regresión, etcétera.

### ¿Qué hace especial a los datos de supervivencia?

Como se mencionó en el apartado anterior, el análisis de supervivencia forma parte de los métodos para analizar estudios de cohorte; entonces ¿qué características de los datos de supervivencia hacen de ésta un área especial de análisis estadístico? En primer lugar, el tiempo de supervivencia es una variable aleatoria positiva, generalmente con sesgo positivo (una cola larga a la derecha), debido a que pocos individuos sobreviven largo tiempo, comparados con la mayoría. Estas características hacen inapropiado el uso de la distribución normal para modelar esta información. Esta dificultad podría superarse mediante la aplicación de alguna transformación a los datos para obtener simetría; sin embargo, es más conveniente trabajar con una distribución alter-

nativa que respete su escala original. El tipo de distribuciones en el análisis de supervivencia es el apropiado para modelar datos con esta peculiaridad.

Una de las características más importante de los datos de supervivencia es la presencia de un fenómeno denominado censura. En los estudios de cohorte puede ocurrir que algún individuo los abandone antes de que le haya ocurrido la falla, por lo que sólo se tendrá información parcial (observación censurada) sobre su tiempo de falla. El objetivo principal de los métodos de supervivencia es incorporar al análisis esta información parcial sobre el tiempo de supervivencia que reportan los individuos censurados. Es conveniente mencionar que en otras áreas estadísticas no se toma en cuenta esta información parcial, debido a que se le considera como "datos faltantes". Esta manera de proceder es contraria a la filosofía estadística de incorporar toda la información disponible dentro del análisis. La presencia de datos censurados es el mayor problema técnico en el desarrollo de métodos estadísticos dentro del análisis de supervivencia, tanto que, incluso métodos tan simples como el análisis descriptivo, requieren procesos más complejos que los habituales.

**Tipos de censura**

La censura en los estudios de supervivencia puede presentarse de diversas maneras, por lo que recibe distintas clasificaciones:



*Censura tipo I*: en muchos estudios de seguimiento, el investigador se ve en la necesidad de fijar un tiempo máximo de observación de los individuos para que les ocurra la falla, usualmente, debido a razones de presupuesto o de diseño del estudio. En este caso, los individuos que al término de este periodo no hayan presentado la falla, constituyen las observaciones censuradas.

*Censura tipo II*: en este caso, el investigador decide prolongar la observación de los individuos en estudio hasta que ocurran $k$ fallas de $n$ posibles ($k \leq n$). Una razón común para determinar el número de fallas que deben observarse es la potencia requerida en el estudio. Los individuos que no experimentaron la falla al completarse estas primeras $k$, representan observaciones censuradas. En estos dos casos, la censura está controlada por el investigador.

*Censura aleatoria*: este tipo de censura ocurre sin que el investigador ejerza ningún control sobre la misma. Las censuras pueden ocurrir porque el individuo abandona el estudio y se pierde su seguimiento, o muere por alguna causa que no está relacionada con el evento de interés.

En estos tres tipos de censura, lo que se sabe es que, de haber ocurrido la falla, ésta se presentaría después del tiempo de censura observado, por lo que se conoce como *censura por la derecha*. Este tipo de censura es el más común en estudios epidemiológicos.

El objetivo de este capítulo es mostrar el uso de las diversas técnicas del análisis de supervivencia a través de un estudio de mujeres con cáncer cérvicouterino (CaCu), que se llevó a cabo con 378 mujeres del Hospital de Ginecobstetricia, número 4 Luis Castelazo Ayala, del Instituto Mexicano del Seguro Social, en la Ciudad de México, que fueron seleccionadas analizando sus registros históricos. Ingresaron a la muestra aquellas mujeres que cumplían con los criterios de inclusión del estudio. Con tal propósito, se definieron las siguientes variables al inicio del estudio:

*Tiempo de supervivencia* (`tiemdd`): se determinó como el tiempo transcurrido desde el momento del diagnóstico de la paciente con CaCu confirmado histopatológicamente (evento inicial), hasta la ocurrencia de la muerte (evento final). Las censuras (`censura1`) ocurrieron debido a mujeres que permanecieron con vida al finalizar el periodo de estudio, murieron por alguna otra causa no relacionada con CaCu o se perdió su seguimiento (cambio de domicilio sin notificación, negativa a continuar dentro del estudio o falta de seguimiento del tratamiento).

*Covariables*: Además de registrar el tiempo de supervivencia de las pacientes se hizo, al inicio del estudio, el registro de algunos factores (covariables) relacionados con el transcurso de la enfermedad y que son potenciales modificadores de este tiempo de supervivencia. Se recolectó información sobre la realización del examen de Papanicolaou (`pap`), síntomas (`síntoma`), tamaño del tumor (`tamaño`) y tipo histológico (`tipo`), entre otras.

## MÉTODOS PARA EL ANÁLISIS DE SUPERVIVENCIA

### Funciones involucradas en el análisis de supervivencia T3

El análisis del tiempo de supervivencia de una población se realiza, principalmente, mediante dos funciones:

1. La función de supervivencia, $S(t)$, que se define como la probabilidad de que la variable $T$ (tiempo de supervivencia) sobrepase un tiempo fijo ($t$). Expresada mediante símbolos, se tiene la siguiente fórmula:

$$S(t) = P(T > t)$$

   que proporciona la probabilidad de que un individuo dentro de la población continúe con vida después de determinado tiempo $t$.

2. La función de riesgo, $h(t)$, que es la tasa instantánea de falla al tiempo $t$.* Esta función se expresa matemáticamente como:

$$h(t) = \underset{\Delta t \to 0}{Lim} \frac{P(t \le T < t + \dot{A}t \,|\, T \ge t)}{\Delta t}$$

Dado que $h(t)$ es una tasa, entonces se sabe que $h(t) \geq 0$, es decir, es positiva, pero no necesariamente acotada.

Esta función es útil para, a partir de ella, proponer algún modelo que se ajuste a los datos. Ya que es común que un investigador posea información relevante sobre el comportamiento de determinado riesgo en una población, esta información puede auxiliarlo para elegir el modelo matemático cuya función de riesgo se ajuste de mejor manera a estas características por él conocidas.

* Una presentación más ilustrativa de esta función se encuentra en la referencia 1.

### Modelos paramétricos

Como en muchas áreas de análisis estadístico, en el análisis de supervivencia existen modelos paramétricos (distribuciones de probabilidad) para modelar los tiempos de supervivencia observados en alguna población. Los modelos más comunes son:

*Modelo Exponencial*: permite modelar fenómenos cuyo riesgo es constante a lo largo del periodo de observación.

*Modelo Weibull*: modela tiempos de supervivencia cuyo riesgo es monótono creciente o decreciente. Es el modelo más común en estudios epidemiológicos.

*Modelo Log-normal*: es útil para modelar individuos cuyo riesgo al inicio del estudio es nulo, crece hasta un valor máximo y después decrece a cero para valores grandes de *t*. No es común que el riesgo en una población se anule cuando *t* crece desmedidamente; por esta razón, este modelo puede ser inadecuado como modelo para el tiempo de supervivencia; sin embargo, es apropiado cuando no se está interesado en valores grandes de *t*.

*Modelo Gamma*: modela riesgos monótonos crecientes que inician en cero y se estabilizan cuando *t* es grande. También se usa cuando los riesgos son inicialmente muy grandes y luego decrecen de manera monótona hasta volverse constantes para valores grandes de *t*. A diferencia del modelo Log-normal, el riesgo no desaparece cuando *t* crece, sino que siempre está latente, y lo hace un modelo más plausible como modelo de supervivencia.



La especificación completa de estos modelos requiere del conocimiento de sus parámetros (de ahí el nombre de modelos paramétricos), mismos que se estiman a partir de los datos disponibles. Posteriormente puede describirse la supervivencia de la población bajo estudio, utilizando el modelo con los parámetros estimados. Este capítulo no ilustra el procedimiento para ajustar un modelo paramétrico a los tiempos de supervivencia de las mujeres con CaCu, ya que, por lo regular, es preferible utilizar el enfoque no paramétrico (que se presenta más adelante). Una ventaja adicional de este último es que puede proporcionar indicios para el ajuste de alguno de los modelos paramétricos mencionados anteriormente.

La figura 1 muestra el comportamiento de algunas de las funciones de riesgo para estos cuatro modelos.

### Métodos no paramétricos

En muchas ocasiones, el investigador desconoce el modelo paramétrico que puede ajustarse a sus datos. En este caso, puede recurrir a métodos de análisis que no supongan una distribución de los datos, conocidos genéricamente como *métodos no paramétricos*. Los más usuales en supervivencia son:

*Tabla de vida*: es el resumen de la información recabada, y es parecida a la tabla de frecuencias usual. Contiene estadísticas que permiten tener una visión del comportamiento de la mortalidad, la supervivencia y la tasa de riesgo en cada intervalo de tiempo. Al igual que en los histogramas, uno de los problemas de este método es la elección del número de intervalos y, en consecuencia, el tamaño de los mismos, para representar la información.

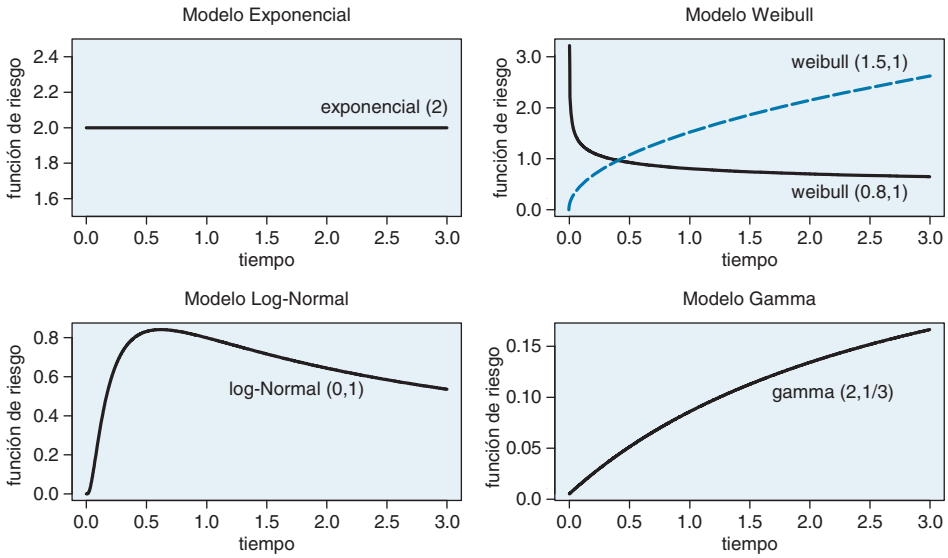



**Figura 1** **Funciones de riesgo para los principales modelos paramétricos en el análisis de supervivencia**



Para obtener esta tabla en STATA se procede de la siguiente forma:* La variable `censura1` está codificada como 1, si el individuo presentó la falla (en este caso, si la paciente murió), y 0 si se censuró.

a. al empezar la sección se define el tiempo de supervivencia y la censura.

```
stset tiemdd censura1
failure event:  censura1 != 0 & censura1 < .
obs. time interval:  (0, tiempo2]
 exit on or before:  failure
------------------------------------------------------------------------------
      378  total obs.
        0  exclusions
------------------------------------------------------------------------------
      378  obs. remaining, representing
      105  failures in single record/single failure data
     1304  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        13
```

* Como en los capítulos anteriores, en adelante se destacan en "negritas" las palabras que refieren comandos específicos del paquete STATA.

b. El cuadro I se obtiene como resultado de unir la información de los comandos **sts list** y **ltable**:

```
1. sts list
2. ltable tiemdd censura1,hazard interval (0,365,730,1095,1460,1825,
   2190,2555,2920,3650,4015,4380,4745)
```

Estos últimos valores indican que las funciones de supervivencia y riesgo se han calculado en intervalos anuales. Si desean calcularse en intervalos distintos, sólo hay que especificar los valores correspondientes después del comando.

El cuadro I muestra la tabla de vida de las mujeres con CaCu. En ella se observa que la mayoría de las muertes por este padecimiento ocurren en los primeros tres años de observación (92 de 107), lo que da como resultado que la probabilidad de sobrevivir a esos tres años se reduzca en 28%, (1-0.72), que es muy alta comparada con la reducción total 56.54%, (1-0.4346) (en términos relativos es aproximadamente 50%). Por supuesto, lo anterior se ve reforzado con el hecho de que en este periodo se tienen las tasas de riesgo, en general, más elevadas.

*El estimador Kaplan-Meier*: un mejor estimador (debido a que se construye con la información desagregada, es decir, con cada dato observado) de la función de supervivencia que el que pro-



**Cuadro I Tabla de vida de las mujeres con cáncer cervicouterino**

| *Intervalo (años)* | *Número de pacientes en el intervalo* | *Número de censuras* | *Número de pacientes expuestas al riesgo* | *Fallas* | *Proporción de fallas* | *Super-vivencia acumulada* | *Función de riesgo* |
|---|---|---|---|---|---|---|---|
| 0-1 | 378 | 27 | 364.5 | **59** | .1619 | .8381 | .0005 |
| 1-2 | 292 | 62 | 261.0 | **19** | .0728 | .7771 | .0002 |
| 2-3 | 211 | 41 | 190.5 | **14** | .0735 | **.7200** | .0002 |
| 3-4 | 156 | 49 | 131.5 | 3 | .0228 | .7036 | .0001 |
| 4-5 | 104 | 33 | 87.5 | 5 | .0571 | .6634 | .0002 |
| 5-6 | 66 | 25 | 53.5 | 2 | .0374 | .6386 | .0001 |
| 6-7 | 39 | 15 | 31.5 | 0 | .0000 | .6386 | .0000 |
| 7-8 | 24 | 7 | 20.5 | 1 | .0488 | .6074 | .0001 |
| 8-9 | 16 | 7 | 12.5 | 1 | .0800 | .5588 | .0002 |
| 9-10 | 8 | 3 | 6.5 | 0 | .0000 | .5588 | .0000 |
| 10-11 | 5 | 1 | 4.5 | 1 | .2222 | .4346 | .0007 |
| 11-12 | 3 | 1 | 2.5 | 0 | .0000 | .4346 | .0000 |
| 12-13 | 2 | 2 | 1.0 | 0 | .0000 | **.4346** | .0000 |

porciona la tabla de vida se obtiene por medio del estimador Kaplan-Meier (*K-M*). Este es, por excelencia, el estimador usual de la función de supervivencia de una población. El uso del *K-M* permite estimar, en cada tiempo *t*, la probabilidad que tiene un individuo en la población de sobrevivir a la falla; además proporciona una manera gráfica de representar esta supervivencia. La figura 2 muestra la gráfica del estimador *K-M* de las mujeres con CaCu. Para determinar cuál es el valor correspondiente de supervivencia en cada tiempo, basta con trazar una línea vertical desde el eje horizontal (tiempo en días) hasta la gráfica y de ahí una línea horizontal hasta el valor de la supervivencia en el eje vertical. En STATA se obtiene con el comando:

```
sts graph
```

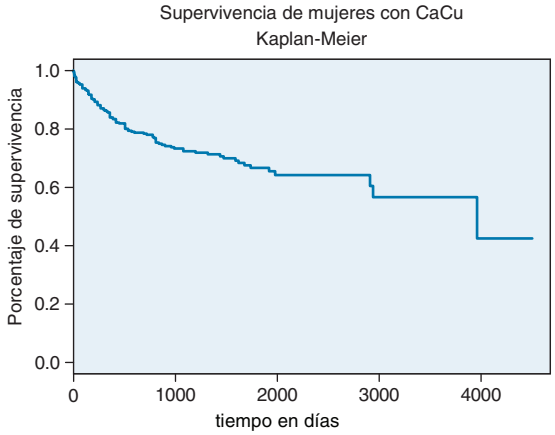



**Figura 2** **Estimador Kaplan-Meier para el tiempo de supervivencia en mujeres con CaCu**

### Comparación de poblaciones

En este tipo de estudios es de gran importancia evaluar el efecto que puede tener alguna característica (covariable) particular de la población sobre la supervivencia. Una manera empírica de explorar si la supervivencia de los individuos es diferente por alguna de sus características, es graficar simultáneamente los estimadores *K-M* por cada una de sus categorías. Obviamente, esta exploración gráfica podría dar indicios de que la supervivencia de una población es distinta por niveles de una covariable; sin embargo, no representa una prueba formal de esta diferencia. El siguiente comando de STATA genera la gráfica para explorar la diferencia en supervivencia entre las mujeres que se hicieron el examen de Papanicolaou y las que no.

```
sts graph,by (pap)
```

Como puede observarse en la figura 3, es evidente que la supervivencia de las mujeres que no se realizaron el examen de Papanicolaou es menor que las que sí se lo realizaron, hecho que pue-

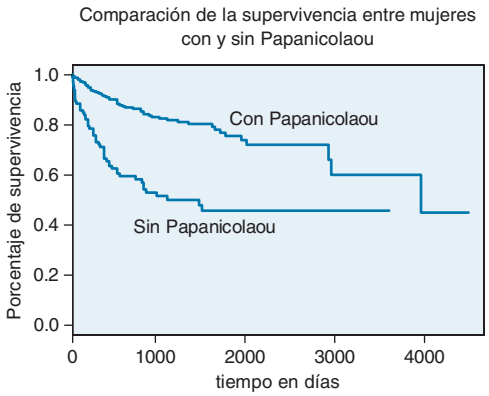



**Figura 3** **Estimadores Kaplan-Meier para mujeres con y sin examen de Papanicolaou**

de observarse por una caída más rápida de la curva de supervivencia de las primeras. No obstante, esto no constituye ninguna prueba formal de esta diferencia.



### Pruebas estadísticas de comparación de poblaciones

Dos de las pruebas no paramétricas más comunes para realizar de manera formal esta comparación son la *log-rank* y la de *Wilcoxon*. En cada una de ellas, la hipótesis nula es que las supervivencias de todas las poblaciones son iguales, mientras que la hipótesis alternativa establece que, por lo menos, una de ellas es distinta, es decir:

$$H_0: S_1 = S_2 = ... = S_k \quad vs. \quad H_1: S_i \neq S_j \text{ para algún } i \neq j \quad i, j = 1,2,...,k$$

con $S_i$ la función de supervivencia en la población $i$.

La diferencia esencial entre estas dos pruebas radica en que la de Wilcoxon es una prueba ponderada. Se recomienda usar log-rank si las poblaciones a comparar presentan riesgos proporcionales (ver más adelante en este mismo capítulo). El cuadro II muestra los valores de las pruebas log-rank y sus correspondientes niveles de significancia, al comparar las poblaciones que definen los diferentes niveles de cada covariable incluida. Como puede observarse, esta prueba confirma la diferencia en la supervivencia entre las mujeres que se hicieron el examen de Papanicolaou y las que no, diferencia que ya se había observado en la figura 3. Se ilustra el uso del comando correspondiente en STATA, con la variable realización del examen de Papanicolaou (pap). De manera similar puede hacerse para el resto de las variables representadas en el cuadro.

```
sts test pap,logrank
```

**Cuadro II Comparación de poblaciones de mujeres con cáncer cervicouterino***

| *Variable* | *Valor de la estadística* | *Valor p* |
|---|---|---|
| Realización del Papanicolaou (pap) | 37.26 | <0.001 |
| Presentación de síntomas (síntoma) | 22.77 | <0.001 |
| Etapa clínica (etapa) | 69.10 | <0.001 |
| Tamaño del tumor en cm (tamaño) | 37.89 | <0.001 |
| Diseminación (disemin) | 30.28 | <0.001 |

* La construcción de las variables involucradas en este cuadro puede consultarse en la referencia 2.

## EL MODELO DE RIESGOS PROPORCIONALES O MODELO DE COX

### Modelos con covariables

En una investigación epidemiológica, es deseable evaluar el efecto conjunto sobre la supervivencia que pueden tener los factores que resulten significativos de manera individual. Uno de los modelos más utilizados para este fin es el modelo de riesgos proporcionales o modelo de Cox. Este es un modelo tipo regresión que especifica cómo cambia la función de riesgo básica o de referencia† respecto de los individuos en la población no básica. Este cambio lo especifica el parámetro asociado a cada covariable introducida al modelo. El modelo de riesgos proporcionales tiene la forma:

$$h(t|x_{j1}, x_{j2}, \ldots, x_{jp}) = h_0(t)\, \exp(\beta_2 x_{j2} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp})$$

donde

$h(t|x_{j1}, x_{j2}, \ldots, x_{jp})$ es el riesgo de un individuo al que se le registraron las covariables $x_{j1}, x_{j2}, \ldots, x_{jp}$, $j = 1,2,\ldots,n$

$h_0(t)$ es la función de riesgo básica, y

$\beta_i$ es el parámetro asociado a la $i$-ésima covariable, $i = 1,2,...,p$, mismo que debe ser estimado.

El nombre de riesgos proporcionales proviene del hecho de que el cociente entre el riesgo de la población básica y la no básica *no depende* de $t$; lo que puede apreciarse al rescribir el modelo como:

$$\frac{h(t|\mathbf{x}_j)}{h_0(t)} = \exp(\beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp})$$

† Esta población no necesariamente existe de manera natural en un estudio de supervivencia, lo que posibilita al investigador para definirla a partir de su conocimiento del problema bajo estudio. Una manera usual de definirla es construirla con los individuos que tienen el nivel basal ("0") en todas sus covariables.

Con $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jp})'$ el vector de covariables asociadas al individuo $j$, $j = 1,2,...,n$.

Obsérvese que el término de la derecha de esta expresión no depende de $t$, es decir, es constante para cualquier valor del tiempo bajo estudio. Esto implica que los riesgos básico y no básico *son proporcionales*. De hecho, este es el supuesto fundamental para ajustar el modelo de Cox, por lo que, en un análisis particular, debe verificarse. Este modelo puede escribirse como un modelo lineal tomando logaritmos en ambos miembros de la igualdad:

$$\ln\left(\frac{h(t|\mathbf{x}_j)}{h_0(t)}\right) = \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp},\ j = 1, 2, \ldots, n$$

**Estimación del modelo**

La estimación del modelo de riesgos proporcionales se realiza por medio de un concepto de verosimilitud desarrollado por David Cox,[3] llamado verosimilitud parcial. Lo atractivo de este tipo de verosimilitud para el modelo de riesgos proporcionales es que puede estimar el efecto de las covariables sobre el tiempo de supervivencia sin necesidad de estimar la función de riesgo básica, dando como resultado una estimación semi-paramétrica del modelo. Por el contrario, si en un análisis del tiempo de supervivencia se especifica la forma funcional (distribución de probabilidad) de la función de riesgo básica, entonces el proceso de estimación resulta ser totalmente paramétrico, ya que se estiman tanto los parámetros involucrados en la función de riesgo básica como los parámetros asociados con las covariables.



Las pruebas de hipótesis sobre la significancia de cada parámetro dentro del modelo se realizan de modo semejante a las del modelo de regresión logística, según se ha visto en el capítulo 14.

**Interpretación de los parámetros**

Consideremos que tenemos un modelo con una única covariable $X$ continua. Supongamos que queremos comparar dos individuos con valores $X = x$ y $X = x + 1$ de esta covariable. Tenemos que $h(t|X = x) = h_0 \exp(\beta x)$ y $h(t|X = x + 1) = h_0 \exp[\beta(x + 1)]$, entonces el logaritmo del cociente de riesgos es $\beta$, por lo que $\hat{\beta}$ se interpreta como el cambio promedio estimado en el logaritmo del cociente de riesgos por unidad de cambio en la covariable $X$. Si se aplica la función exponencial, se tiene que el cociente de riesgos es $\exp(\beta)$ y, por lo tanto, $\exp(\hat{\beta})$ se interpreta como el cambio promedio estimado en el cociente de riesgos. Esta última interpretación es la más común en este modelo.

Si la diferencia entre los valores de estos individuos es ahora $r$, entonces el valor del cociente de riesgos es $\exp(r\beta)$ y su valor estimado se interpreta como el cambio promedio estimado en el cociente de riesgos cuando la covariable se incrementa en $r$ unidades. Es importante notar que cuando la covariable es continua, el cociente de riesgos no depende del valor específico de la covariable, sino de la magnitud del cambio entre los valores de la misma; en otras palabras, si la covariable fuera la edad, el cociente de riesgos de una mujer con edad 50, comparada con una de 55, sería el mismo que el de una de 20, comparada con una de 25, hecho no necesariamente plausible en una investigación epidemiológica. Una manera de subsanar esta anomalía es introducir un factor cuyos niveles correspondan a diferentes conjuntos de valores de la covariable, es decir, categorizarla.

Cuando la covariable tiene *j* niveles, se requiere un nivel de comparación que, por lo general, es el nivel base o "0" de la covariable. Si denotamos por $\gamma_j$ el parámetro asociado al nivel *j* de la covariable, entonces $\exp(\gamma_j)$ es el cociente de riesgos entre una mujer de la población base y una del nivel *j* de esta covariable; por lo tanto, su valor estimado se interpreta como el cambio promedio estimado en el cociente de riesgos entre ellas.

Los factores (covariables) que modifican de manera importante la supervivencia de las mujeres con CaCu, se reportan en el cuadro III. Para implementar el modelo, se generaron las variables indicadoras de cada categoría. Las instrucciones para correr este modelo en STATA son:

```
stcox pap eta2 eta3 eta4  sint2 sint3 hist1 hist3
```

**Cuadro III Factores asociados a la supervivencia de las mujeres con cáncer cervicouterino**

| | *MODELO DE RIESGOS PROPORCIONALES* | | |
|---|---|---|---|
| *VARIABLE* | *RR*** | *Valor p* | *IC95%* |
| Realización del Papanicolaou | | | |
| No | 1.00* | | |
| Sí | 0.46 | <0.001 | 0.30-0.70 |
| Etapa clínica | | | |
| I | 1.00* | | |
| II | 1.02 | 0.924 | 0.57-1.85 |
| III | 1.93 | 0.025 | 1.08-3.44 |
| IV | 5.99 | <0.001 | 2.83-12.67 |
| Presentación de síntomas | | | |
| Síntomas generales | 1.00* | | |
| Síntomas específicos | 2.50 | 0.054 | 0.98-6.36 |
| Síntomas graves | 3.26 | 0.019 | 1.21-8.72 |
| Tipo histológico | | | |
| Adenocarcinoma | 1.00* | | |
| Epidermoide | 1.38 | 0.296 | 0.75-2.52 |
| Adenoescamoso | 2.24 | 0.036 | 1.05-4.78 |

* Categoría de referencia
** Razón de riesgos del modelo de riesgos proporcionales, ajustada por las variables incluidas en el modelo.

Puede observarse que el riesgo de morir por este padecimiento es, en promedio, 50% menor (*RR*=0.46) en las mujeres que se hicieron el examen de Papanicolaou, en relación con las que no se lo hicieron (población de referencia), una vez que se ha ajustado por las demás covariables. Los parámetros estimados para el resto de los factores se interpretan de manera semejante.

## EVALUACIÓN DEL AJUSTE DEL MODELO DE RIESGOS PROPORCIONALES

### Distintos tipos de residuos para evaluar el ajuste del modelo

En el análisis de supervivencia existen diversos residuos[4] que se usan para verificar algunas de las características del modelo de riesgos proporcionales. Los residuos de Cox-Snell sirven para evaluar lo adecuado del ajuste del modelo; los de Martingalas para verificar si es adecuado incluir de manera lineal las covariables en el modelo; los residuos de devianza se utilizan para detectar residuos grandes, que podrían ser potenciales *valores extremos*; la base para las pruebas formales sobre el supuesto de riesgos proporcionales se obtiene por medio de los residuos de Schoenfeld y transformaciones de los residuos de *score* permiten evaluar el impacto que tiene cada observación en el modelo, tanto de manera global (en todo el vector de parámetros estimados) como por cada covariable.

### Riesgos proporcionales



El supuesto más importante del modelo de Cox es que los riesgos entre la población base y la población no base son proporcionales. Este supuesto puede verificarse para cada covariable incluida en el modelo o de manera global. Existen alternativas gráficas que no constituyen una prueba formal, y alternativas basadas en estadísticos de prueba. La primera de éstas consiste en graficar alguna transformación del estimador *K-M* de cada población contra una transformación del tiempo de supervivencia. Si los riesgos son proporcionales, la gráfica debería mostrar líneas aproximadamente paralelas entre las poblaciones.

```
stphplot,by (pap)
```

Las pruebas estadísticas se construyen a partir de los residuos de Schoenfeld. En estas pruebas, la hipótesis nula establece que los riesgos de las poblaciones son proporcionales contra la hipótesis de que, por lo menos, una población no presenta un riesgo proporcional a las restantes.

La figura 4 muestra que es razonable suponer que se cumple el supuesto de riesgos proporcionales para la variable *realización del examen de Papanicolaou.* Por otro lado, en el cuadro IV se verifica que todas las variables incluidas en el modelo cumplen este supuesto, excepto para la variable *síntomas específicos.* Esta podría introducirse como una variable dependiente del tiempo (ver más adelante en este mismo capítulo) o realizarse un análisis estratificado con ella.[1,5] También puede verse que globalmente se cumple este supuesto.

Para generar los residuos de Schoenfeld se corre nuevamente el modelo con las siguientes instrucciones:

```
a) stcox pap eta2 eta3 eta4  sint2 sint3  hist1 hist3, schoenfeld(sc*) scaledsch(ssc*)
b) stphtest, rank detail
```

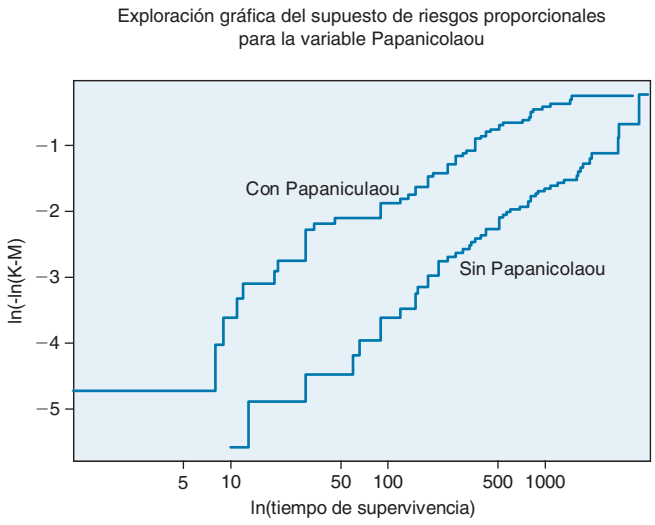



**Figura 4** **Supuesto de riesgos proporcionales**



**Cuadro IV** **Evaluación global y por covariable del supuesto de riesgos proporcionales**

| *Variable* | *Valor* p* |
|---|---|
| Realización del examen de Papanicolaou | 0.0578 |
| Síntomas específicos | 0.0278 |
| Síntomas graves | 0.0595 |
| Cáncer tipo epidermoide | 0.5989 |
| Cáncer tipo adenoescamoso | 0.9071 |
| Etapa II | 0.2983 |
| Etapa III | 0.2258 |
| Etapa IV | 0.8081 |
| Evaluación global | 0.1595 |

* Prueba de *ji* cuadrada.

### Ajuste del modelo

Como se mencionó anteriormente, para verificar lo adecuado del ajuste del modelo de riesgos proporcionales se utilizan los residuos de Cox-Snell. Si el modelo de riesgos proporcionales se ajusta adecuadamente a los datos, estos residuos deben tener una distribución exponencial con parámetro igual a *uno*. En consecuencia, si graficamos una transformación del estimador *K-M* de estos residuos contra ellos mismos, debemos observar una línea recta con pendiente unitaria.

Para generar la gráfica correspondiente deben realizarse los pasos que se enlistan a continuación:

a. **stcox** pap eta2 eta3 eta4 sint2 sint3 hist1 hist3,**mgale(mg)** (se corre nuevamente el modelo con el comando mgale)
b. **predict** cs,**csnell** (se construyen los residuos de Cox-Snell)
c. **stset** cs censura1 (se construye el estimador *K-M* a partir de estos residuos)
d. **sts** gen H=s [se genera una nueva variable($H$) con los valores del *K-M* del paso anterior]
e. **gen** H1=-ln(H) (se realiza una transformación (-l$n$($H$)) de la variable generada)
f. **scatter** H1 cs (se realiza la gráfica entre la variable transformada y los valores de los residuos de Cox-Snell)

La figura 5 muestra que existen algunas observaciones que el modelo no ajusta adecuadamente, ya que están separadas de la recta a 45°. No obstante, podemos considerar que, en general, se obtiene un buen ajuste del modelo de riesgos proporcionales para esta población. Nótese



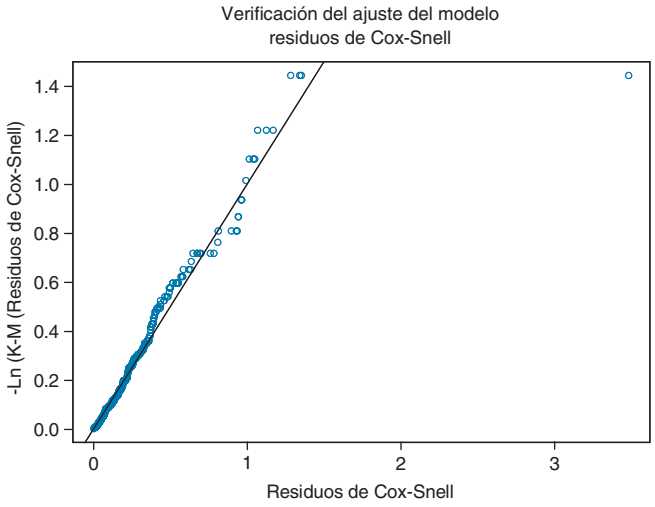



**Figura 5** **Residuos de Cox-Snell**

la presencia de una observación totalmente alejada de la línea recta, misma que es una potencial observación aberrante.

**Evaluación de otras características del modelo**

Si la gráfica de los residuos de martingalas contra los valores de cada covariable no muestra ningún patrón aparente, es decir, los puntos en la gráfica están dispersos en forma aleatoria, entonces puede considerarse que es adecuada la inclusión de la covariable de forma lineal en el modelo.

La gráfica de los residuos de devianza* contra el tiempo debe mostrar también un patrón aleatorio, con los puntos dispuestos de manera relativamente simétrica alrededor del eje horizontal. Además, pueden observarse residuos grandes, indicativos de probables valores extremos. El comando de STATA para obtener estos residuos es:

```
quietly predict double dev,deviance

ksm dev tiemdd
```

En la figura 6 se aprecian residuos grandes de individuos con baja supervivencia. Se observa que estos residuos se distribuyen de manera aproximadamente simétrica alrededor del cero, por lo que se considera que el modelo se ajusta adecuadamente a los datos. No se muestra ninguna gráfica para los residuos de martingalas, porque todas las covariables incluidas en el modelo son nominales y, en este caso, estas gráficas no son muy ilustrativas.



* Los residuos de martingalas no son simétricos alrededor del cero, sino que tienen un fuerte sesgo positivo. Los residuos de devianza son una transformación de los de martingalas con la finalidad de tener residuos que sean aproximadamente simétricos alrededor del cero.

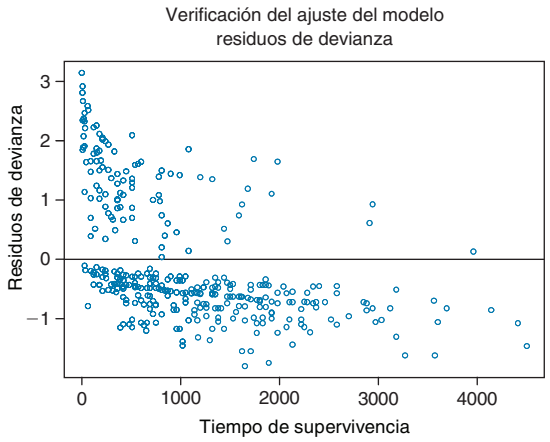



**Figura 6** **Residuos de devianza**

### Diagnóstico del modelo

Un paso muy importante en la evaluación del ajuste de un modelo consiste en verificar el impacto de las observaciones en sus parámetros estimados. Una observación puede tener un fuerte impacto en la estimación *de todo el vector de parámetros*, o bien, en *alguno de los parámetros* en particular. Para cuantificar la magnitud de este impacto en el modelo, se utilizan transformaciones de los residuos score.

En el primer caso, se tiene un solo valor (*dfbeta*) que especifica el cambio observado en la estimación de todo el vector de parámetros cuando la observación es removida; si este cambio es grande, puede decirse que la observación es influyente en la estimación del vector de parámetros. En el caso de una evaluación por variable, cada individuo tiene tantos valores como covariables contenga el modelo (*dfbetas*). Cada uno de estos valores representa el cambio observado en la estimación del parámetro asociado con la covariable en cuestión, cuando la paciente ha sido removida. Si la magnitud de este cambio es grande, puede afirmarse que la observación es influyente en la estimación del parámetro correspondiente.

Primeramente, se describirá el proceso en STATA para obtener las gráficas para verificar el impacto que cada observación tiene sobre cada una de las variables incluidas en el modelo.

1. Correr el modelo definiendo los residuos

   **`stcox`** `pap eta2 eta3 eta4 sint2 sint3 hist1 hist3,` **`esr(esr*)`** (se corre el modelo y se generan los residuos de `score (esr(esr*))`



2. Generación de matrices

   **`set matsize 400`** (ampliación de la memoria)

   **`mkmat`** `esr1 esr2 esr3 esr4 esr5 esr6 esr7 esr8,` **`matrix(esr)`** (se construye una matriz con los vectores de residuos score generados)

   **`mat`** `V=e(V)` (se define una matriz(V) con la matriz de varianzas y covarianzas de los estimadores [e(V)])

   **`mat`** `Inf=esr*V` [se define una matriz con el producto de las dos matrices generadas anteriormente (se escalan los residuos de score)]

   **`svmat`** `Inf,names(s)` [toma las columnas de la matriz Inf y las guarda como variables (*s*: *s*1, *s*2,...,*s*8). Cada una de estas variables representa los residuos escalados de score de cada variable dentro del modelo, en el orden en que introdujeron al modelo: `pap, eta2`, etcétera.]

   Enseguida, se procede a graficar cada una de estas variables(s) contra el tiempo de supervivencia. La línea continua que aparece en las gráficas se genera con el comando **`ksm`**.

El proceso para generar en STATA la gráfica correspondiente a la evaluación global del impacto de cada observación en todo el modelo es el siguiente:

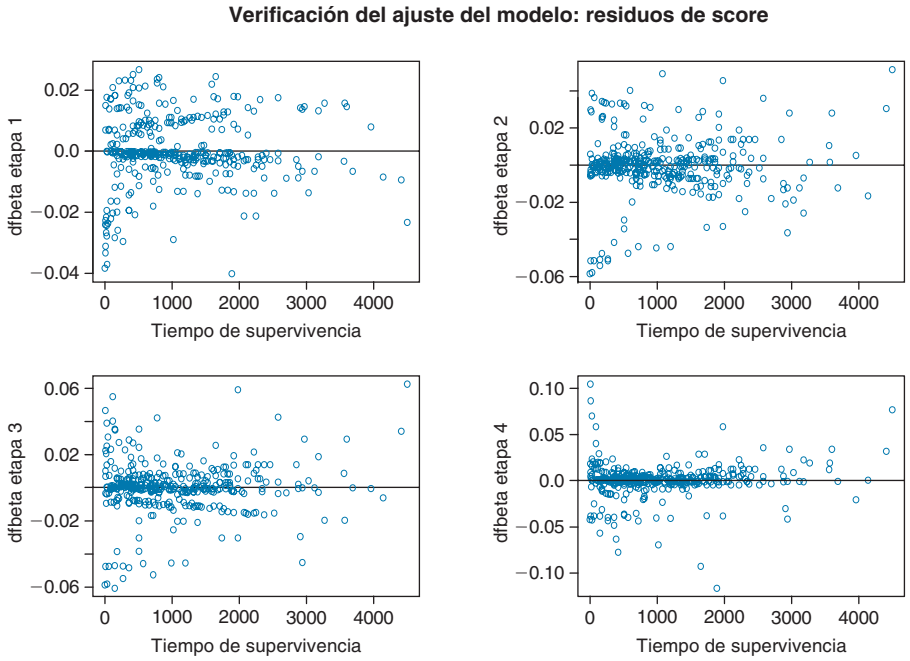



**Figura 7a** **Residuos de score para evaluar el impacto de cada covariable**



**mat** `B=Inf*esr'` [se construye una matriz (B) como el producto de la matriz de residuos escalados de score (Inf) y la matriz transpuesta (') de los residuos de score (esr)]

**matrix symeigen** `x lambda=B` (se calcula la matriz de eigenvectores de B)

**svmat** `double x` [se toman las columnas de esta matriz y se guardan como variables(x)]

**gen** `x1a=abs(x1)` [se genera una variable, que es el valor absoluto de la primera variable(`x1a`) generada en el proceso anterior]. Posteriormente se grafica esta variable contra el tiempo de supervivencia.

Sólo una observación muestra un fuerte impacto en todo el vector de parámetros estimados, como lo muestra la figura 8. El valor de *dfbeta* para esta paciente es de 0.7680914 (esta estadística se estandariza de manera que la suma de sus cuadrados sea uno[6] y la aportación de esta observación es de 0.5899; casi 60%). Las características que hacen atípica a esta paciente son que pertenece a una etapa tardía del padecimiento (etapa IV) y su registro de supervivencia es grande comparado con el promedio del resto de observaciones de su etapa (1650 días vs. 347 días

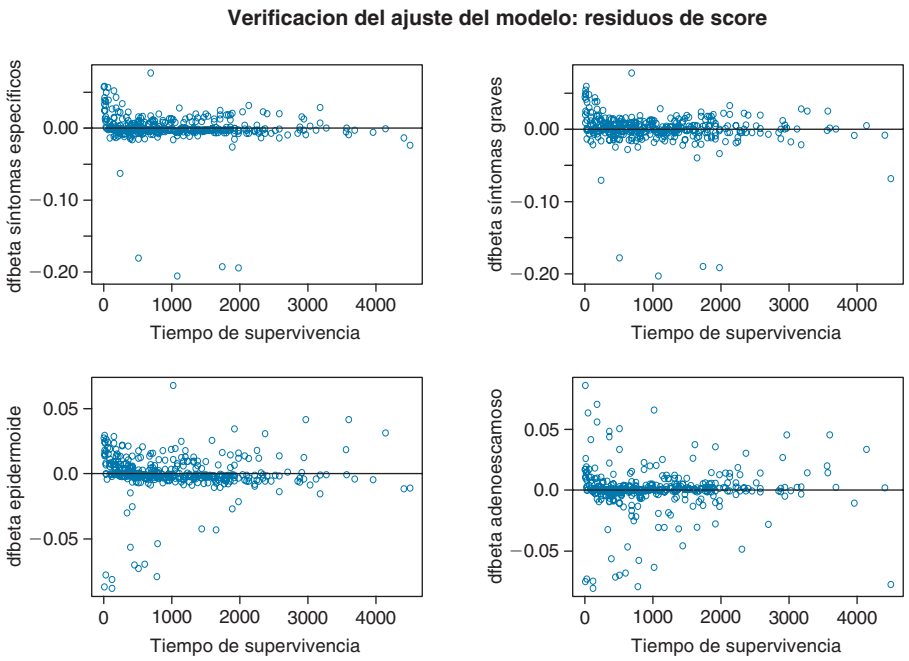



**Figura 7b** **Residuos de score para evaluar el impacto de cada covariable**

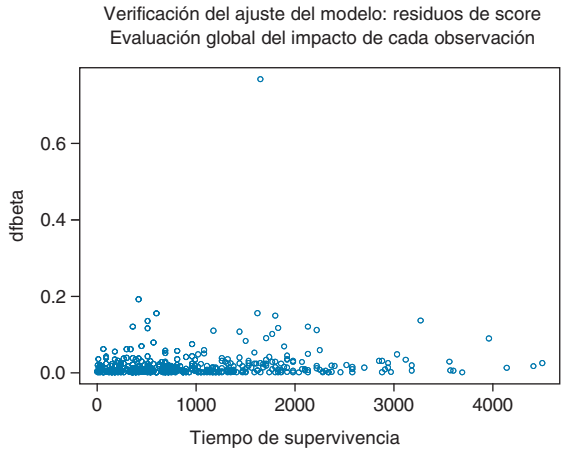



**Figura 8** **Residuos de score para evaluar el impacto de una observación en el vector de parámetros**

promedio); además, representa una observación censurada, lo que implica que, por lo menos, sobrevivió 1650 días. Al remover esta observación y correr el modelo sin la misma, sólo hubo un cambio notorio en el coeficiente asociado con la etapa IV (5.9965 *vs*. 8.1095). En la evaluación del impacto de las observaciones sobre cada una de las covariables introducidas al modelo (*dfbetas*, figuras 7a y 7b), se encontró que el impacto más fuerte lo ejerce esta misma observación sobre la variable etapa IV, lo que confirma el resultado obtenido con el *dfbeta*.

## OTROS TIPOS DE ESTUDIOS DE SUPERVIVENCIA

Las técnicas presentadas hasta ahora sirven para analizar estudios de supervivencia que pueden considerarse como estándar; sin embargo, existen algunas características especiales que requieren el uso de otro tipo de técnicas. En los siguientes párrafos presentaremos algunas de estas características y mencionaremos las técnicas de uso común en los estudios.

### Extensión del modelo de Cox

Existen diversas situaciones en los estudios epidemiológicos en los que se tienen covariables que dependen del tiempo; una de las más comunes es la aplicación de un medicamento. La dosis de un medicamento que se suministra para contrarrestar los efectos de una enfermedad depende de su grado de severidad a lo largo del estudio. Entonces, si se evalúa el efecto del medicamento sobre la supervivencia de los individuos, y dado que la dosis del mismo puede variar en el transcurso del análisis, es lógico considerar esta covariable como dependiente del tiempo. En este tipo de situaciones, el modelo estándar de riesgos proporcionales no es adecuado, ya que los riesgos entre las poblaciones no serían proporcionales; en este caso, se utiliza una extensión de este modelo que permite la introducción de estas covariables dependientes del tiempo. Este modelo es semejante al anterior y sólo incorpora el término $x_j(t)$ en lugar de $x_j$ en aquellas covariables que dependan del tiempo. Aunque este cambio en la notación es simple, la estructura de la base de datos y la estimación del modelo son mucho más complejos que en el modelo estándar.

### Eventos múltiples por individuo

Existen dos situaciones consideradas como eventos múltiples de un individuo,[7] cuando los eventos son del mismo tipo, por ejemplo: infartos agudos de miocardio, varias crisis diabéticas, etc., y eventos de distinta naturaleza: varios tipos de falla. Se han propuesto diversos modelos para analizar este tipo de estudios, entre los que se encuentran los llamados modelos marginales y los modelos con efectos aleatorios, conocidos como *frailty* en la bibliografía sobre el tema de supervivencia. Las anteriores son sólo algunas situaciones donde los modelos de supervivencia básicos no dan una respuesta satisfactoria. Por supuesto, existen muchos otros tipos de estudio que requieren de métodos ad hoc para su análisis. El lector interesado puede consultar las referencias bibliográficas de este capítulo.

Los análisis presentados en este capítulo se realizaron en el paquete STATA 8.0 y las gráficas se hicieron con el paquete S-PLUS 2000.

## REFERENCIAS


1. Kleinbaum DG. Survival analysis. A self-learning text. Nueva York: Springer-Verlag, 1996.
2. Flores-Luna L, Zamora-Muñoz S, Salazar-Martínez E, Lazcano-Ponce E. Análisis de supervivencia. Aplicación en una muestra de mujeres con cáncer cervical en México. Salud Publica Mex 2000;42(3):242-251.
3. Cox DR. Partial likelihood. Biométrica 1975; 62:269-276.
4. Therneau TM, Grambsch PM, FlemingTR. Martingale-based residuals for survival models. Biometrika 1990;77(1):147-160.
5. Hougaard P. Fundamentals of survival data. Biometrics 1999;55:13-22.
6. Collett D. Modeling survival data in medical research. Nueva York: Chapman & Hall, 1994.
7. Therneau TM, Grambsch PM. Modeling survival data. Extending the Cox model. Nueva York: Springer-Verlag, 2000.

# XVI Interacción o modificación de efecto en modelos de regresión lineal y logística*

*Héctor Lamadrid Figueroa*
*Martha María Téllez Rojo Solís*

## INTRODUCCIÓN

La modificación de efecto, llamada también interacción, es uno de los conceptos más importantes en epidemiología. Este concepto describe un fenómeno que se observa cuando la magnitud de un efecto, es decir, el estimador de la asociación entre la exposición y el evento de interés, se modifica al cambiar los valores de una tercera variable. Esta característica distingue un modificador de efecto de un confusor, el cual, además, se encuentra necesariamente asociado con el evento de interés y la exposición.[1] Las bases del concepto de modificación de efecto fueron establecidas por Miettinen[1] y su relación con la noción de multicausalidad se discutió en manuscritos posteriores.[2,3]

En teoría, la modificación de efecto, a diferencia de la interacción, establece una jerarquía de relaciones causales. Los términos sinergismo (o antagonismo) se usan a veces como sinónimos de interacción o modificación de efecto,[4] aunque éstos han de usarse con cautela y conocimiento del modelo epidemiológico subyacente.[5] Aun cuando existan las diferencias conceptuales mencionadas, en esta sección se usarán indistintamente los términos modificación de efecto e interacción.

La detección de interacción estadística no necesariamente arroja luz sobre la naturaleza de un proceso biológico.[6] Sin embargo, la estimación de interacciones es indispensable, por lo menos, por tres motivos: 1) permite la estimación del efecto combinado de dos o más variables; 2) mejora la capacidad predictiva del modelo estadístico, y 3) permite enfocar medidas preventivas a grupos con mayor riesgo.[7]

La estimación de la modificación del efecto en un estudio epidemiológico es relativamente fácil de interpretar cuando la variable modificadora es dicotómica o categórica, y se observa la heterogeneidad de la medida de efecto entre los estratos del modificador.[8] Sin embargo la interpretación de modelos de regresión que incluyen términos de interacción con variables continuas o discretas adquiere mayor complejidad y no ha sido ampliamente discutida.

El objetivo de esta sección es la revisión de algunos conceptos básicos de interacción, haciendo énfasis en la interpretación de modelos multivariados de regresión lineal y logística que incluyen modificación de efecto por variables continuas o discretas, y sugiere una manera de representar gráficamente la presencia de interacción. Aunque sólo se incluyen ejemplos de los

* Agradecemos al Dr. Eduardo César Lazcano Ponce y la M. en C. Alma Ethelia López Caudana por permitir el uso de las bases de datos utilizadas en los ejemplos de este capítulo.

métodos estadísticos mencionados, los principios aquí formulados pueden aplicarse a cualquier tipo de modelo lineal generalizado. En este capítulo no se discute la estimación de los intervalos de confianza de las medidas de interacción, que puede revisarse en cualquiera de los textos especializados.[8-12]

## INTERACCIÓN ADITIVA Y MULTIPLICATIVA

Existen dos formas de interacción, a saber: aditiva y multiplicativa. La primera se define como "alejamiento de la aditividad de los efectos en una escala particular de medición de resultados".[13] Si dos factores $X_1$ y $X_2$ influyen sobre una variable resultado en forma independiente uno del otro, es de esperarse que el efecto combinado de ambos factores sea igual a la suma de sus efectos individuales. De cumplirse esto, estaríamos en presencia de una "aditividad de efectos" o estado de no-interacción. Por el contrario, si los factores interactúan entre ellos, su efecto combinado es mayor (o menor) a la suma de los efectos individuales, y existe un "alejamiento de la aditividad" o estado de interacción. La interacción multiplicativa, en cambio, se define como el "alejamiento de la multiplicatividad de los efectos", es decir, existe una interacción cuando la multiplicación de los efectos individuales de dos o más factores no es igual a su efecto combinado.

Aunque durante un tiempo existió cierta controversia respecto a cuál de estas dos formas de evaluar la presencia de interacción es mejor,[6,14,15] ahora se acepta que esto depende de la medida de efecto que se utilice, y ésta, a su vez, del contexto del problema (estadístico, biológico, salud pública, etc.). De hecho, como señala Rothman,[13] la presencia o ausencia de interacción depende estrictamente del modelo estadístico que se utilice. En general, la interacción se evalúa como un alejamiento de la aditividad en modelos en los que el efecto se mide como una diferencia (regresión lineal, diferencias de riesgos, etc.) y como un alejamiento de la multiplicatividad en modelos en que el efecto se mide como una razón (razón de momios, riesgo relativo, etcétera).



En presencia de una interacción, el efecto de una exposición sobre la probabilidad de ocurrencia de una enfermedad o sobre la magnitud de una variable de respuesta deja de ser una constante para convertirse en una función de la variable modificadora del efecto de la exposición. Dicho efecto, por lo tanto, es diferente para cada valor de la variable modificadora, lo cual concuerda con la definición de Miettinen.[1]

Supongamos que deseamos evaluar el efecto de dos variables independientes ($X_1$ y $X_2$) sobre una variable dependiente continua $Y$ (sin distinguir cuál de las dos juega el papel de una exposición principal). El *efecto* o asociación* de X sobre Y se define como la diferencia en el valor de la media de $Y$ al comparar dos poblaciones que difieren en una unidad de una variable independiente X. En el caso de un estudio longitudinal, esta diferencia implica un cambio, pero esto también es aplicable en un estudio transversal donde la comparación válida es entre subpoblaciones que difieren en una unidad consecutiva.

* La diferencia entre el término efecto o asociación depende del diseño del estudio epidemiológico. Para poder hablar del primero requerimos necesariamente de un estudio longitudinal.

La siguiente ecuación representa un modelo estimado donde $\hat{Y}$ es la respuesta esperada y $X_1$ y $X_2$ denotan variables independientes que pueden ser continuas, dicotómicas o discretas. El tercer término de la ecuación corresponde al *término de interacción*.*

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\delta}_1 X_1 X_2$$

Para evaluar el efecto de cada variable independiente se verá qué es lo que sucedería si se incrementara su valor en una unidad. Se define $y_0$ como el valor inicial promedio de $Y$ y $y_1$ como el valor de la media de $Y$ al incrementar $X_1$ en una unidad:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_1 x_1 x_2$$

$$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + 1) + \hat{\beta}_2 x_2 + \hat{\beta}_1 (x_1 + 1) x_2$$

Sea $\hat{\Delta} y_{x1}$ el efecto de $X_1$ (la diferencia promedio entre $Y$ con $X_1$ y la correspondiente con $X_1 + 1$):

$$\hat{\Delta} y_{x_1} = \hat{y}_1 - \hat{y}_0 = \hat{\beta}_1 + \hat{\delta}_1 x_2$$

siguiendo el mismo razonamiento, sea $\hat{\Delta} y_{x_2}$ el efecto de $X_2$:

$$\hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_2 + 1) + \hat{\delta}_1 x_1 (x_2 + 1)$$

$$\hat{\Delta} y_{x_2} = \hat{y}_2 - \hat{y}_0 = \hat{\beta}_2 + \hat{\delta}_1 x_1$$

por lo tanto, la suma de los efectos individuales† de ambas variables es:

$$\hat{\Delta} y_{x_1} + \hat{\Delta} y_{x_2} = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2) \qquad (1)$$

Evalúese ahora cuál sería el efecto combinado de ambas variables, incrementando *simultáneamente* su valor en una unidad:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\delta}_1 x_1 x_2$$

$$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + 1) + \hat{\beta}_2 (x_2 + 1) + \hat{\delta}_1 (x_1 + 1)(x_2 + 1)$$

* Para evaluar la presencia de interacción en el modelo se incluye una nueva variable que resulta de la multiplicación de las variables que interactúan, el coeficiente de esta nueva variable es sometido entonces a prueba de hipótesis.

† A veces llamados "efectos marginales".

Donde $y_1$ es el valor promedio de $Y$ cuando $X_1$ y $X_2$ se incrementan simultáneamente en una unidad. El efecto combinado ($\hat{\Delta}y_{x_1x_2}$) de ambas variables es entonces:

$$\hat{\Delta}y_{x_1x_2} = \hat{y}_1 - \hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 x_1 + \hat{\delta}_1 x_2 + \hat{\delta}_1$$

$$\hat{\Delta}y_{x_1x_2} = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1(x_1 + x_2 + 1) \quad (2)$$

Existe aditividad de efectos (no interacción), sólo si el efecto combinado de ambas variables es igual a la suma de sus efectos individuales, o sea:

$$\hat{\Delta}y_{x_1x_2} = \hat{\Delta}y_{x_1} + \hat{\Delta}y_{x_2}$$

$$\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1(x_1 + x_2 + 1) = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1(x_1 + x_2) \quad (3)$$

La ecuación 3 es verdadera si y sólo si $\hat{\delta}_1$ (el coeficiente de regresión lineal del término de interacción) es igual a 0. De hecho:

$$\hat{\Delta}y_{x_1x_2} - (\hat{\Delta}y_{x_1} + \hat{\Delta}y_{x_2}) = \hat{\delta}_1 \quad (4)$$

El coeficiente del término de interacción $\hat{\delta}_1$ es la diferencia entre el efecto combinado de ambas variables independientes y la suma de sus efectos individuales. Existe interacción si y sólo si $\hat{\delta}_1$ es diferente de 0. Por esta razón, para probar la existencia de interacción, tendrá que hacerse la prueba de hipótesis: $H_0$: $\delta_1 = 0$ *vs.* $H_1$: $\delta_1 \neq 0$.



En un modelo de regresión logística se evalúa la presencia de interacción multiplicativa si se utilizan como medidas de efecto las razones de momios (*RM*). Existe una interacción si el efecto combinado de ambos factores es diferente que la multiplicación de sus efectos individuales. A continuación, se muestra un modelo estimado de regresión logística, en el que se asume la presencia de una interacción y una relación lineal entre la variable modificadora de efecto y el logit de *p*:

$$\text{logit}\,\hat{p} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\delta}_1 X_1 X_2$$

El efecto individual de $X_1$ sobre el logit de *p* se estima de la siguiente manera:

$$\text{logit}\,\hat{p}_1 = \hat{\beta}_0 + \hat{\beta}_1(x_1 + 1) + \hat{\beta}_2 x_2 + \hat{\delta}_1(x_1 + 1)x_2$$

$$\text{logit}\,\hat{p}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\delta}_1 x_1 x_2$$

$$\text{logit}\,\hat{p}_1 - \text{logit}\,\hat{p}_0 = \hat{\beta}_1 + \hat{\delta}_1 x_2$$

El efecto sobre el logit de *p* puede expresarse también como *RM*. Sea $\hat{\Psi}_{X_1}$ la *RM* estimada al comparar los momios del evento en poblaciones con $X_1$ +1 y $X_1$, $\frac{\hat{p}_1}{1-\hat{p}_1}$ es el momio estimado para la población con $X_1$ + 1 y $\frac{\hat{p}_0}{1-\hat{p}_0}$ es el momio estimado para la población con $X_1$:

$$\ln\left(\frac{\hat{p}_1}{1-\hat{p}_1}\right) - \ln\left(\frac{\hat{p}_0}{1-\hat{p}_0}\right) = \hat{\beta}_1 + \hat{\delta}_1 x_2$$

Por propiedades de los logaritmos esto es equivalente a:

$$\ln\left(\frac{\frac{\hat{p}_1}{1-\hat{p}_1}}{\frac{\hat{p}_0}{1-\hat{p}_0}}\right) = \hat{\beta}_1 + \hat{\delta}_1 x_2$$

$$\hat{\Psi}_{X_1} = \frac{\frac{\hat{p}_1}{1-\hat{p}_1}}{\frac{\hat{p}_0}{1-\hat{p}_0}} = e^{\hat{\beta}_1 + \hat{\delta}_1 x_2}$$

En forma semejante la *RM* de $X_2$ es:

$$\hat{\Psi}_{X_2} = e^{\hat{\beta}_2 + \hat{\delta}_1 x_1}$$

por lo tanto, la multiplicación de las *RM* individuales de ambas variables es:

$$\hat{\Psi}_{X_1}\hat{\Psi}_{X_2} = e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2)} \tag{5}$$

La *RM* combinada de ambas variables es:

$$\hat{\Psi}_{X_1 X_2} = e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2 + 1)} \tag{6}$$

Existe multiplicatividad de efectos (no interacción), sólo si el efecto combinado de ambas variables es igual a la multiplicación de sus efectos individuales, o sea:

$$\hat{\Psi}_{X_1 X_2} = \hat{\Psi}_{X_1}\hat{\Psi}_{X_2}$$

$$e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2 + 1)} = e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2)} \tag{7}$$

por lo tanto, la multiplicatividad se cumple si y sólo si $\hat{\delta}_1$ (el coeficiente de regresión logística del término de interacción) es igual a 0. De hecho:

$$\frac{e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2 + 1)}}{e^{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\delta}_1 (x_1 + x_2)}} = e^{\hat{\delta}_1} \tag{8}$$

Existe interacción si y sólo si $\hat{\delta}_1$ es diferente de 0; por lo tanto, para probar la existencia de interacción en regresión logística pueden tenerse las siguientes pruebas de hipótesis: $H_0$: $\delta_1 = 0$ *vs.* $H_1$: $\delta_1 \neq 0$, o bien, $H_0$: $e^{\delta_1} = 1$ *vs.* $H_1$: $e^{\delta_1} \neq 1$, donde $e^{\hat{\delta}_1}$ es lo que erróneamente se ha llamado "RM" del término de interacción (ver más adelante).

Cuando se utiliza la *RM* como medida de efecto, generalmente se evalúa la presencia de interacción multiplicativa. Sin embargo, si se utiliza el logaritmo natural de las razones de momios (es decir los coeficientes de la regresión logística), se observa que se está tratando, en realidad, con interacción aditiva.

$$\ln \hat{\Psi}_{X_1X_2} = \ln \hat{\Psi}_{X_1} \hat{\Psi}_{X_2}$$
$$\ln(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\delta}_1(x_1+x_2+1)}) = \ln(e^{\hat{\beta}_1+\hat{\beta}_2+\hat{\delta}_1(x_1+x_2)})$$
$$\hat{\beta}_1+\hat{\beta}_2+\hat{\delta}_1(x_1+x_2+1) = \hat{\beta}_1+\hat{\beta}_2+\hat{\delta}_1(x_1+x_2)$$

Nótese que esta última ecuación es exactamente igual a la ecuación (3). Es evidente que la naturaleza aditiva o multiplicativa de la interacción está completamente determinada por la escala en la que se están midiendo los efectos. Finalmente, como la ausencia de una interacción estadísticamente significativa en una escala particular no implica la ausencia de una interacción significativa en alguna otra escala,[13] es recomendable especificar en qué medida de efecto se encuentra o no la presencia de una interacción.

## MODIFICACIÓN DE EFECTO EN REGRESIÓN LINEAL



Supóngase que se tiene un modelo estimado de regresión lineal donde, en la media de la variable de respuesta para un cierto valor de X y M, *X* denota la variable de exposición (continua, discreta o binaria) y *M* es una variable modificadora de efecto continua o discreta.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 M + \hat{\delta}_1 XM$$

$\hat{\Delta} y_X$ denota el efecto de *X*, que es una función de la variable modificadora *M*:

$$\hat{\Delta} y_X(M) = \hat{\beta}_1 + \hat{\delta}_1 M \qquad (9)$$

donde $\hat{\beta}_1$ es el coeficiente de regresión asociado a la exposición *X* y $\hat{\delta}_1$ es el coeficiente de regresión asociado con el término de interacción. La ecuación (9) puede utilizarse como fórmula para obtener el efecto de la exposición sobre *Y* en presencia de una modificadora de efecto.

Cuando *M* se incrementa en saltos de una unidad, el efecto de *X* sobre *Y* en presencia de una variable modificadora de efecto *M*, constituye una sucesión aritmética, cuya diferencia común es el coeficiente del término de interacción $\hat{\delta}_1$. El coeficiente $\hat{\delta}_1$ es, por lo tanto, la diferencia en el efecto de la exposición entre unidades sucesivas de la variable modificadora. Véanse algunos ejemplos.

**Ejemplo 1**

El cuadro I muestra los resultados obtenidos de un estudio transversal realizado en 1 491 mujeres trabajadoras del IMSS en Cuernavaca, Morelos. El objetivo de este estudio fue estimar la aso-

**Cuadro I** **Coeficientes y valores *p* de un modelo estimado de regresión lineal para evaluar el efecto del número de embarazos sobre la DMO (*n* = 1 491)**

| *Variable* | *Coeficiente* | *p* |
|---|---|---|
| Número de embarazos | 23.48 | <0.01 |
| Edad | −1.95 | <0.01 |
| NE × ED | −0.47 | <0.01 |
| Constante | 510.53 | <0.01 |

*NE* × *ED* = interacción número de embarazos por edad

ciación entre el número de embarazos (NE) y la densidad mineral ósea (DMO), en presencia de la edad (ED) como modificadora de efecto. Se propuso el siguiente modelo:

$$\hat{DMO} = \hat{\beta}_0 + \hat{\beta}_1 NE + \hat{\beta}_2 ED + \hat{\delta}_1 NE \times ED$$

El coeficiente del término de interacción es significativo ($p < 0.01$), por lo que hay evidencia de que la asociación entre el NE y la DMO depende de la edad. En consecuencia, debe estimarse dicha asociación sustituyendo diferentes edades en la ecuación (9):

$$\hat{\Delta} DMO_{NE}(ED) = 23.48 + (-0.47) ED$$

Por ejemplo, para conocer cuál es la asociación entre el número de embarazos y la DMO en una mujer de 20 años, se sustituye en la fórmula:

$$\hat{\Delta} DMO_{NE}(20) = 23.48 + (-0.47)(20)$$
$$\hat{\Delta} DMO_{NE}(20) = 14.16$$

**Cuadro II** **Asociación estimada entre el número de embarazos y la DMO en mujeres de 25 a 30 años**

| *Edad (años)* | *Efecto del número de embarazos (mg/cm²/embarazo)* |
|---|---|
| 25 | 11.82 |
| 26 | 11.35 |
| 27 | 10.89 |
| 28 | 10.42 |
| 29 | 9.95 |
| 30 | 9.49 |

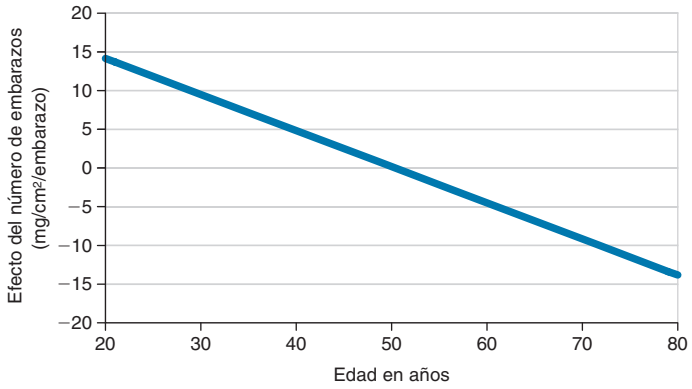



**Figura 1** **Asociación entre el número de embarazos y la DMO según la edad**



De acuerdo con el modelo propuesto, se estima que, en promedio, en las mujeres de 20 años de edad, la DMO es mayor en 14.16 mg/cm$^2$ en comparación con la población de aquellas que han tenido un embarazo menos.

El cuadro II muestra la asociación estimada para las edades de 25 a 30 años. Se observa que la diferencia en la magnitud de la asociación entre años sucesivos es constante: −0.47; exactamente el valor del coeficiente del término de interacción. Este valor es la pendiente de la recta que surge al graficar $\hat{\Delta}DMO_{NE}(ED)$ (figura 1). En la gráfica se observa cómo la asociación estimada entre el número de embarazos y la *DMO* disminuye linealmente con el incremento de la edad.

En conclusión, existe evidencia que sugiere que la asociación entre el número de embarazos y la DMO va disminuyendo conforme aumenta la edad, y que la disminución se mantiene a un ritmo constante a través de los años (0.47 mg/cm$^2$ por embarazo).

## MODIFICACIÓN DE EFECTO EN REGRESIÓN LOGÍSTICA

El siguiente es un modelo de regresión logística, donde $p$ es la probabilidad de ocurrencia de una enfermedad, $X$ denota la variable de exposición (continua, discreta o binaria) y $M$ es una variable modificadora de efecto. El modelo estimado es:

$$\text{logit}\,\hat{p} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 M + \hat{\delta}_1 XM$$

$\hat{\Psi}_X$ denota la *RM* de $X$, que es una función de la variable modificadora $M$:

$$\hat{\Psi}_X(M) = e^{\hat{\beta}_1 + \hat{\delta}_1 M}$$

Esto es equivalente a:

$$\hat{\Psi}_X(M) = e^{\hat{\beta}_1}(e^{\hat{\delta}_1})^M \tag{10}$$

donde $e^{\hat{\beta}_1}$ es la razón de momios cuando $M = 0$. A esta cantidad se le llamará RM basal, denotándola como $\hat{\Psi}_{X0}$. Sea $\hat{\lambda}_1$ la exponencial del coeficiente del término de interacción. De tal manera que si: $e^{\hat{\delta}_1} = \hat{\lambda}_1$ y $e^{\hat{\beta}_1} = \hat{\Psi}_{X0}$, entonces puede volverse a escribir la expresión (10) como:

$$\hat{\Psi}_X(M) = \hat{\Psi}_{X0}\hat{\lambda}_1{}^M \tag{11}$$

La ecuación (11) puede utilizarse para obtener la *RM* de cualquier exposición en presencia de una variable modificadora de efecto.

Si se obtuvieran las *RM* de la asociación entre la exposición y el evento para valores sucesivos de la variable modificadora ($m$ y $m + 1$; donde m es un entero) y posteriormente obtuviésemos la razón entre ambas medidas, el resultado sería justamente $\hat{\lambda}_1$ [ecuación (12)]. Este valor, mal llamado "RM" de la interacción, es en realidad *la razón de las* RM *de la exposición entre unidades sucesivas de la variable modificadora de efecto*:

$$\hat{\lambda}_1 = \frac{\hat{\Psi}_X(m+1)}{\hat{\Psi}_X(m)} \tag{12}$$



**Ejemplo 2**

En el cuadro III se presentan los resultados de un estudio de casos y controles en 682 mujeres mexicanas. El modelo de regresión logística propuesto pretende estimar la asociación entre el antecedente de embarazos (*AE*) y el cáncer mamario, con la edad (*ED*) como modificadora de efecto. Su forma estimada es:

$$\text{logit}\,\hat{p} = \hat{\beta}_0 + \hat{\beta}_1 AE + \hat{\beta}_2 ED + \hat{\delta}_1 AE \times ED$$

Los resultados muestran evidencia de que la *RM* asociada al antecedente de embarazos previos es 1.045 veces mayor por cada año que se incrementa la edad. La modificación de efecto es

**Cuadro III Razones de momios y valores *p* de un modelo estimado de regresión logística para evaluar el efecto del antecedente de embarazos sobre el cáncer mamario (*n* = 682)**

| *Variable* | *Razón de momios* | *p* |
|---|---|---|
| Antecedente de embarazo | 0.026 | 0.001 |
| Edad | 0.985 | 0.342 |
| AE × ED | 1.045 | 0.024 |

*AExED* = interacción antecedente de embarazo por edad.

**Cuadro IV** **Razón de momios de la asociación entre antecedente de embarazo y cáncer mamario en mujeres de 50 a 55 años**

| *Edad (años)* | *Razón de momios* |
|---|---|
| 50 | 0.238 |
| 51 | 0.249 |
| 52 | 0.260 |
| 53 | 0.272 |
| 54 | 0.284 |
| 55 | 0.297 |

significativa ($p = 0.024$), por lo que debemos estimar el efecto del antecedente de embarazos para cada edad, sustituyendo en la ecuación (11):

$$\hat{\Psi}_{AE}(ED) = (0.026)(1.045)^{ED}$$

Por ejemplo, la *RM* del antecedente de embarazo, cuando la edad es igual a 23 años, es:

$$\hat{\Psi}_{AE}(23) = 0.026(1.045)^{23} = 0.071$$



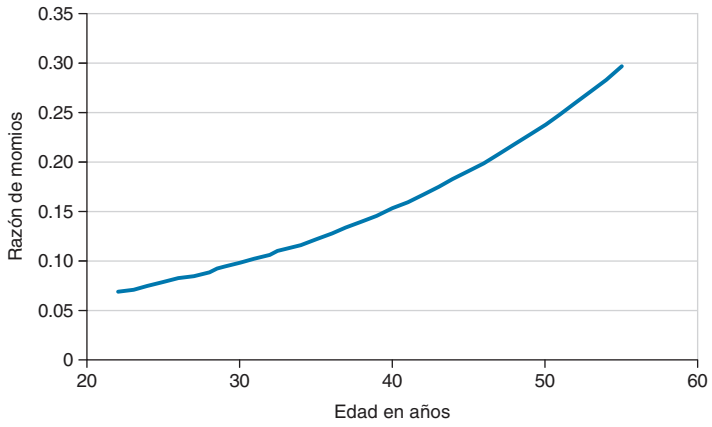



**Figura 2** **Razón de momios de la asociación entre el antecedente de embarazos y el cáncer mamario según la edad**

Las *RM* de la asociación entre el antecedente de embarazos previos y cáncer mamario en las edades de 50 a 55 años se presentan en el cuadro IV. Se observa que la razón entre cualquier *RM* y la correspondiente al año inmediato anterior es constante: 1.045: el valor de la "*RM*" del término de interacción $\hat{\lambda}_1$. La *RM* aumenta exponencialmente al incrementarse la edad (figura 2).

## EVALUACIÓN DE LA INTERACCIÓN CON STATA

En STATA puede utilizarse el comando **lincom** para encontrar el efecto de la exposición sobre la variable de interés, tomando en cuenta la presencia de modificación de efecto. Retomando el ejemplo 1 del texto, y suponiendo que interesa estimar la asociación entre el número de embarazos y la DMO en mujeres de 40 años. El primer paso es generar el modelo de regresión lineal propuesto:

```
regress dmo ne ed neXed
```

```
  Source |       SS       df       MS              Number of obs =    1491
---------+------------------------------           F(  3,  1487) =  135.03
   Model |  1307986.68     3   435995.56           Prob > F      =  0.0000
Residual |  4801275.62  1487  3228.83364           R-squared     =  0.2141
---------+------------------------------           Adj R-squared =  0.2125
   Total |  6109262.31  1490  4100.17604           Root MSE      =  56.823

------------------------------------------------------------------------------
     dmo |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ne |   23.47715   3.533588      6.644   0.000       16.54581     30.4085
      ed |  -1.946093   .2392665     -8.134   0.000      -2.415428   -1.476757
  neXed  |  -.4663057    .072868     -6.399   0.000      -.6092408   -.3233707
   _cons |   510.5291   10.17851     50.158   0.000       490.5633    530.4949
------------------------------------------------------------------------------
```

A continuación, se utiliza el comando **lincom**, sumando el coeficiente de la exposición con el coeficiente de término de interacción, multiplicado por el valor específico de la variable modificadora de interés (`ed=40`):

```
lincom ne+40*neXed
```

```
 ( 1)  ne + 40.0 neXed = 0.0
------------------------------------------------------------------------------
     dmo |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     (1) |   4.824923   1.084082      4.451   0.000        2.69843    6.951415
------------------------------------------------------------------------------
```

En mujeres de 40 años la DMO es, en promedio, 4.82 mg/cm$^2$ menor en comparación con aquellas que han tenido un embarazo menos. Con el comando **lincom** tenemos la ventaja adicional de obtener el intervalo de confianza del efecto de la exposición para este valor de la variable modificadora.

Ahora, utilizando los datos del ejemplo 2, supóngase que nos interesa conocer cuál es la asociación entre el antecedente de embarazo y el cáncer mamario (`cam`) en una mujer de 55 años de edad. Al generar el modelo de regresión logística propuesto, se tiene:

```
logistic cam ae ed aeXed

Logit estimates                                   Number of obs   =        682
                                                  LR chi2(3)      =      23.10
                                                  Prob > chi2     =     0.0000
Log likelihood =  -244.9227                       Pseudo R2       =     0.0450

------------------------------------------------------------------------------
         cam | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ae |   .0257508   .0289033    -3.260   0.001       .0028535    .2323793
          ed |   .9848906    .015769    -0.951   0.342       .9544639    1.016287
       aeXed |    1.04546   .0206017     2.256   0.024       1.005851    1.086629
------------------------------------------------------------------------------
```



El comando **`lincom`** utiliza los coeficientes de la regresión logística y no las RM, por esta razón se le pide una suma, y no una multiplicación. El resultado se expresa en RM, si se utilizó con anterioridad el comando **`logistic`** para generar el modelo.

```
. lincom ae + 55*aeXed

 ( 1)  ae + 55.0 aeXed = 0.0

------------------------------------------------------------------------------
         cam | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .2969607   .0859148    -4.197   0.000       .1684359    .5235561
------------------------------------------------------------------------------
```

Debido a que la *RM* es menor de la unidad, el antecedente de embarazo resulta ser un factor protector contra el cáncer mamario. Los resultados se interpretan señalando que una mujer de 55 años con antecedente de embarazo tiene 3.37 (1/0.297) veces menos posibilidades de tener cáncer mamario que una mujer de la misma edad sin dicho antecedente.

## CONCLUSIONES

Se ha visto que la presencia de interacción o modificación de efecto está definida por la heterogeneidad de la medida de asociación entre dos variables según los valores de una tercera. Los coeficientes de los términos de interacción representan los cambios que se observan en la magnitud de la asociación entre dos variables, cuando se transita entre los distintos estratos de una variable que modifica dicha relación. Esta relación es aditiva (diferencia de coeficientes) en el caso de modelos de regresión lineal y multiplicativa (razón de razones de momios) en modelos de regresión logística. Los ejemplos manejados en esta sección se hicieron con variables modificadoras continuas o discretas, puesto que la interacción con variables dicotómicas es en realidad un caso particular de las continuas, en la cual la variable modificadora sólo puede asumir dos valores.

Las fórmulas presentadas para la estimación de los efectos o asociaciones (ecuaciones 10 y 11) pueden generalizarse para estimar asociaciones cuando existe una interacción de la exposición con más de una variable. Interacciones de orden superior, en las cuales las variables modificadoras interactúan entre sí, también pueden evaluarse en modelos de regresión lineal y logística; sin embargo esto va más allá del alcance de esta sección, por lo que se remite al lector a textos especializados.[10,11]

## REFERENCIAS


1. Miettinen O. Confounding and effect-modification. Am J epidemiol 1974;100:350-353.
2. Rothman K, Causes. Am J Epidemiol 1976;104:587-592.
3. Koopman J, Causal models and sources of interaction. Am J epidemiol 1977;106:439-444.
4. Gordis L. Epidemiology, 2a. ed. Filadelfia: W. B. Saunders Co., 2000
5. Darroch J. Biologic sinergism and paralellism. Am J Epidemiol 1997;145:661-668.
6. Weed D, Selmon M, Sinks T. Links between categories of interaction. Am J Epidemiol 1988;127:17-27.
7. Thompson WD, Effect modification and the limits of biological inference from epidemiologic data. J Clin Epidemiol 1991;44:221-232.
8. Kupper L, Hogan M. Interaction in epidemiologic studies. Am J Epidemiol 1978;108:447-453.
9. Figueiras A, Domenech-Massons JM, Cadarso C. Regression models: calculating the confidence interval of effects in the presence of interactions. Stat Med 1998;17:2099-2105.
10. Kleinbaum, David G. Logistic regression: a self-learning text. Nueva York: Springer-Verlag, 1994:142-149.
11. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. Boston: PWS-Kent Publishing Co., 1988:163-180.
12. Hosmer DW, Lemeshow S. Applied logistic regression. Nueva York: Wiley, 1989:104.
13. Rothman K, Greenland S. Modern Epidemiology. 2a. ed. Philadelphia: Lippincot-Raven, 1998.
14. Rothman K, Greenland S, Walker A. Concepts of interaction. Am J Epidemiol 1980;112:467-470.
15. Pearce N. Analytical implication of epidemiological concepts of interaction. Int J Epidemiol 1989;18:976-980.