

Laura Moreno Altamirano

Tercera edición

# Epidemiología Clínica

[booksmedicos.org](http://booksmedicos.org)

Mc  
Graw  
Hill



# Epidemiología Clínica

**Tercera edición**

## **Laura Moreno Altamirano**

Médica, con estudios de Maestría en Salud Comunitaria,  
Doctora en Antropología.

Profesora Titular C de la Facultad de Medicina  
de la Universidad Nacional Autónoma de México (UNAM).

Profesora de pregrado y profesora y tutora del Programa  
de Maestrías y Doctorados de Ciencias Médicas, Odontológicas  
y de la Salud, UNAM.

Jefa del Departamento de Salud Pública de la Facultad  
de Medicina de la UNAM.

[booksmedicos.org](http://booksmedicos.org)



MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA  
MADRID • NUEVA YORK • SAN JUAN • SANTIAGO • SÃO PAULO  
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI  
SAN FRANCISCO • SIDNEY • SINGAPUR • ST. LOUIS • TORONTO

*Director editorial:* Javier de León Fraga  
*Editor de desarrollo:* Manuel Bernal Pérez  
*Composición y formación:* Foto Grafic & Diseño  
*Diseño de portada:* Pamela González  
*Supervisora de producción:* Ángela Salas Cañada

#### NOTA

La medicina es una ciencia en constante desarrollo. Conforme surjan nuevos conocimientos, se requerirán cambios de la terapéutica. El (los) autor(es) y los editores se han esforzado para que los cuadros de dosificación medicamentosa sean precisos y acordes con lo establecido en la fecha de publicación. Sin embargo, ante los posibles errores humanos y cambios en la medicina, ni los editores ni cualquier otra persona que haya participado en la preparación de la obra garantizan que la información contenida en ella sea precisa o completa, tampoco son responsables de errores u omisiones, ni de los resultados que con dicha información se obtengan. Convendría recurrir a otras fuentes de datos, por ejemplo, y de manera particular, habrá que consultar la hoja informativa que se adjunta con cada medicamento, para tener certeza de que la información de esta obra es precisa y no se han introducido cambios en la dosis recomendada o en las contraindicaciones para su administración. Esto es de particular importancia con respecto a fármacos nuevos o de uso no frecuente. También deberá consultarse a los laboratorios para recabar información sobre los valores normales.

## EPIDEMIOLOGÍA CLÍNICA

Prohibida la reproducción total o parcial de esta obra,  
por cualquier medio, sin autorización escrita del editor.



# Educación

DERECHOS RESERVADOS © 2013, 2005, 1994, respecto a la tercera edición, por  
MCGRAW-HILL INTERAMERICANA EDITORES, S.A. de C.V.

*A subsidiary of The McGraw-Hill Companies, Inc.*

Prolongación Paseo de la Reforma 1015, Torre A, Piso 17, Col. Desarrollo Santa Fe,  
Delegación Álvaro Obregón

C.P. 01376, México, D.F.

Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

**ISBN: 978-607-15-0826-3**

1234567890  
Impreso en México

1245678903  
Printed in Mexico



# Colaboradores

**Fátima del Carmen Aguilar Díaz** Cirujana Dentista y Maestra en Salud Pública.  
Investigadora, Laboratorios de Biológicos y Reactivos de México, BIRMEX.

**Jesús Arias Gómez** Médico y Maestro en Ciencias Médicas.  
Jefatura de la Consulta Externa, Unidad Médica de Alta Especialidad (UMAE), Hospital de Pediatría,  
Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social (IMSS).

**Jorge Carreón García** Médico y Maestro en Ciencias Sociomédicas con énfasis en Epidemiología.  
Director de Desarrollo Operativo, Comisión de Presupuesto de Salud.  
Comisión Nacional de Protección Social en Salud. Secretaría de Salubridad y Asistencia (SSA).

**Patricia Clark** Médica, Reumatóloga. Jefa de la Unidad de Epidemiología Clínica, Hospital  
Infantil de México, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesora, Tutora y Coordinadora del Área de Epidemiología Clínica del Programa de Maestría  
y Doctorado en Ciencias Médicas, Odontológicas y de la Salud, Facultad de Medicina, UNAM.

**Guadalupe S. García de la Torre** Médica y Maestra en Epidemiología.  
Profesora de pregrado del Departamento de Salud Pública, Facultad de Medicina, Universidad  
Nacional Autónoma de México (UNAM).  
Profesora y tutora del Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas  
y de la Salud, Facultad de Medicina, UNAM.  
Coordinadora de Enseñanza del Departamento de Salud Pública, Facultad de Medicina, UNAM.

**Juan José García García** Médico y Maestro en Epidemiología.  
Profesor de pregrado del Departamento de Salud Pública de la Facultad de Medicina, Universidad  
Nacional Autónoma de México (UNAM).

**Horacio García Romero** Médico Cirujano, Doctor en Bioética.  
Profesor de Historia de la Medicina en Filosofía y Catedrático de Bioética, Facultad de Medicina,  
Universidad Nacional Autónoma de México (UNAM) y Escuela Médico Militar.

**Juan Garduño Espinoza** Médico, Especialista en Medicina Interna, Maestro y Doctor  
en Ciencias Médicas.  
Diplomado en Alta Dirección de Empresas en el Instituto Panamericano de Alta Dirección  
de Empresas (IPADE).  
Profesor y Tutor del Programa de Maestría y Doctorado en Ciencias Médicas Odontológicas  
y de la Salud, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesor de Epidemiología Clínica y Medicina Basada en Evidencias del Departamento de Salud  
Pública, Facultad de Medicina, UNAM.  
Profesor en el Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM)  
Subdirector de Investigación en el Hospital Infantil de México “Federico Gómez”.

**Alma Rosa González Montiel** Química Bióloga Parasitóloga, QBP.  
Responsable Sanitaria, Laboratorios de Biológicos y Reactivos de México, BIRMEX.

**María Eugenia Jiménez Corona** Médica, Maestra en Ciencias y Doctora en Epidemiología.  
Responsable del Área de Investigación en los Laboratorios de Biológicos y Reactivos de México,  
BIRMEX.

**Pablo Kuri Morales** Médico, Maestro en Epidemiología. Profesor y Tutor del Programa de Maestría y Doctorado en Ciencias Médicas Odontológicas y de la Salud, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Subsecretario de Prevención y Promoción a la Salud, Secretaría de Salud.

**Saúl León Hernández** Médico y Maestro en Ciencias.  
Jefe de División de Investigación.  
Instituto Nacional de Rehabilitación. SSA

**Luis Limón Limón** Médico y Maestro en Administración Militar.  
Coordinador de Programación Académica.  
Academia Nacional de Medicina de México.

**Sergio López Moreno** Médico, Especialista en Medicina Familiar y Comunitaria y Doctor en Epidemiología.  
Profesor Titular C en la Unidad Xochimilco de la Universidad Autónoma Metropolitana (UAM).

**Alejandra Moreno Altamirano**  
Cirujana Dentista y Maestra en Epidemiología.  
Profesora de pregrado del Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesora y tutora del Programa de Maestría y Doctorado en Ciencias Médicas Odontológicas y de la Salud, Facultad de Medicina, UNAM.

**Mario Enrique Rendón Macías** Médico, Pediatra, Maestro en Ciencias Médicas.  
Investigador de la Unidad de Investigación en Epidemiología Clínica, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social (IMSS).  
Profesor de Epidemiología Clínica y Medicina Basada en Evidencias del Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesor y tutor del programa de Maestrías y Doctorados de Ciencias Médicas, Odontológicas y de la Salud, Facultad de Medicina, UNAM.

**Rodolfo Rivas Ruiz** Médico, Pediatra, Maestro en Epidemiología Clínica.  
Unidad de Epidemiología Clínica, Hospital Infantil de México, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesor del Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud, Facultad de Medicina, UNAM.

**Elizabeth Salazar Rojas** Residente de Dermatología.  
Hospital General “Dr. Manuel Gea González”.

**Adriana Leticia Valdez González** Médica Cirujana, Especialista en Medicina Interna y Endocrinología. Candidata a Maestra en Ciencias Médicas.  
Unidad de Investigación en Epidemiología Clínica, Unidad Médica de Alta Especialidad (UMAE), Hospital de Especialidades, Centro Médico Siglo XXI, Instituto Mexicano del Seguro Social (IMSS).  
Profesora de Epidemiología Clínica y Medicina Basada en Evidencias del Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).

**Florencia Vargas Voráckova (q.e.p.d.)** Médica, Gastroenteróloga.  
Investigadora del Departamento de Gastroenterología del Instituto Nacional de la Nutrición “Salvador Zubirán”.  
Profesora y tutora del programa de Maestrías y Doctorados de Ciencias Médicas, Odontológicas y de la Salud, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).

**Diana Vilar-Compte** Médica, Maestra en Epidemiología.

Médica adscrita al Departamento de Infectología, Instituto Nacional de Cancerología.  
Profesora de Epidemiología Clínica y Medicina Basada en Evidencias, Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Tutora del Programa de Maestría de Ciencias Médicas, Odontológicas y de la Salud, UNAM.

**Antonio Villa Romero** Médico, Maestro en Salud Pública, Maestro en Ciencias en Epidemiología.

Profesor de Epidemiología Clínica y Medicina Basada en Evidencias, Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesor de la Escuela de Medicina de la Universidad Panamericana.  
Profesor y Tutor del Programa de Maestría de Ciencias Médicas, Odontológicas y de la Salud, UNAM.  
Coordinador de Investigación y Posgrado del Departamento de Salud Pública, Facultad de Medicina, UNAM.

**Miguel Ángel Villasís Kever** Médico, Maestro en Ciencias Médicas.

Investigador Titular, Unidad de Epidemiología Clínica, Hospital de Pediatría,  
Centro Médico Siglo XXI, Instituto Mexicano del Seguro Social (IMSS).

**Niels H. Wachter** Médico, Especialista en Medicina Interna y Maestro en Ciencias Médicas.

Jefe de la Unidad de Investigación en Epidemiología Clínica, Unidad Médica de Alta Especialidad (UMAE), Hospital de Especialidades, Centro Médico Siglo XXI, Instituto Mexicano del Seguro Social (IMSS).  
Profesor de Epidemiología Clínica y Medicina Basada en Evidencias, Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM).  
Profesor y Tutor del programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de Ciencias de la Salud, Facultad de Medicina, UNAM.

# Contenido

|                                                                                    |      |
|------------------------------------------------------------------------------------|------|
| <b>Colaboradores</b> .....                                                         | iii  |
| <b>Prefacio</b> .....                                                              | viii |
| <b>Prólogo a la segunda edición</b> .....                                          | xi   |
| <b>Prólogo a la primera edición</b> .....                                          | xiii |
| <b>Agradecimientos</b> .....                                                       | xiv  |
| <b>Capítulo 1.</b> El desarrollo de la Epidemiología Clínica y su método .....     | 1    |
| Laura Moreno Altamirano                                                            |      |
| Sergio López Moreno                                                                |      |
| <b>Capítulo 2.</b> La fundamentación del problema de investigación .....           | 10   |
| Horacio García Romero                                                              |      |
| Pablo Kuri Morales                                                                 |      |
| Saúl León Hernández                                                                |      |
| <b>Capítulo 3.</b> Lineamientos para el diseño del proyecto de investigación ..... | 19   |
| Laura Moreno Altamirano                                                            |      |
| <b>Capítulo 4.</b> Diseños metodológicos en Epidemiología .....                    | 35   |
| Laura Moreno Altamirano                                                            |      |
| Alejandra Moreno Altamirano                                                        |      |
| <b>Capítulo 5.</b> Estudios experimentales. Ensayo clínico aleatorio .....         | 52   |
| Diana Vilar-Compte                                                                 |      |
| Elizabeth Salazar Rojas                                                            |      |
| <b>Capítulo 6.</b> Farmacovigilancia .....                                         | 73   |
| María Eugenia Jiménez Corona                                                       |      |
| Alma Rosa González Montiel                                                         |      |
| Fátima del Carmen Aguilar Díaz                                                     |      |
| <b>Capítulo 7.</b> Bioética y Epidemiología Clínica .....                          | 87   |
| Horacio García Romero                                                              |      |
| Luis Limón Limón                                                                   |      |
| <b>Capítulo 8.</b> Revisiones sistemáticas y metaanálisis en Medicina .....        | 94   |
| Patricia Clark                                                                     |      |
| Rodolfo Rivas Ruiz                                                                 |      |
| <b>Capítulo 9.</b> Clinimetría .....                                               | 109  |
| Niels H. Wachter                                                                   |      |

|                          |                                                              |            |
|--------------------------|--------------------------------------------------------------|------------|
| <b>Capítulo 10.</b>      | <b>Noción de normalidad en Medicina: usos y limitaciones</b> | <b>140</b> |
|                          | Mario Enrique Rendón Macías                                  |            |
| <b>Capítulo 11.</b>      | <b>Concepto de causalidad en Medicina</b>                    | <b>153</b> |
|                          | Florencia Vargas Voráckova (q.e.p.d.)                        |            |
| <b>Capítulo 12.</b>      | <b>Diagnóstico y evaluación de pruebas diagnósticas</b>      | <b>166</b> |
|                          | Laura Moreno Altamirano                                      |            |
| <b>Anexo 1</b>           |                                                              | <b>191</b> |
|                          | C. García Barrios                                            |            |
| <b>Anexo 2</b>           |                                                              | <b>196</b> |
|                          | H. García Romero                                             |            |
| <b>Capítulo 13.</b>      | <b>Estimación del pronóstico de la enfermedad</b>            | <b>198</b> |
|                          | Antonio Villa Romero                                         |            |
|                          | Mario Enrique Rendón Macías                                  |            |
| <b>Capítulo 14.</b>      | <b>Análisis de decisión en la práctica médica</b>            | <b>216</b> |
|                          | Niels H. Wachter                                             |            |
|                          | Leticia Adriana Valdez González                              |            |
| <b>Capítulo 15.</b>      | <b>Calidad de vida: aproximaciones a su medición</b>         | <b>243</b> |
|                          | Miguel Ángel Villasís Keever                                 |            |
|                          | Jesús Arias Gómez                                            |            |
| <b>Capítulo 16.</b>      | <b>Evaluación de la calidad de la atención médica</b>        | <b>256</b> |
|                          | Juan Garduño Espinosa                                        |            |
| <b>Capítulo 17.</b>      | <b>Selección del análisis estadístico</b>                    | <b>275</b> |
|                          | Jorge Carreón García                                         |            |
|                          | Laura Moreno Altamirano                                      |            |
|                          | Guadalupe S. García de la Torre                              |            |
| <b>Capítulo 18.</b>      | <b>Significancia estadística y significancia clínica</b>     | <b>283</b> |
|                          | Juan José García García                                      |            |
| <b>Capítulo 19.</b>      | <b>Muestreo y cálculo de tamaño de muestra</b>               | <b>292</b> |
|                          | Juan José García García                                      |            |
| <b>Índice alfabético</b> |                                                              | <b>309</b> |



# Prefacio

En este libro, la tercera edición de *Epidemiología Clínica*, más de 25 profesionales; epidemiólogos y epidemiólogos clínicos, nos dimos a la tarea de transmitir nuestra experiencia a fin de elaborar un texto que permita comprender las aplicaciones más actuales de esta disciplina. En la formación de un espíritu científico, la Epidemiología Clínica se concibe como una estrategia de categorización de los fenómenos que se dan en la enfermedad humana, a fin de llegar a conclusiones válidas por medio de la investigación, o bien, con base en la lectura crítica y sistematizada de la información existente en la literatura médica, para así tomar las decisiones más correctas en la práctica médica.

Es importante enfatizar que fue John R. Paul quien en 1938 utilizó por primera vez el término “Epidemiología Clínica”, al dictar la conferencia “Una nueva filosofía para viejas enfermedades”, en la *American Society for Clinical Investigation*, a fin de proponer la utilización de la Epidemiología en el área clínica. Sin embargo, fue hasta fines del decenio de 1960-1969 y principios del de 1970-1979 cuando se incorporó en realidad el término “Epidemiología Clínica”. Feinstein fue, sin duda, quien lo retomó e impulsó vigorosamente, y afirmó que la Epidemiología Clínica era una disciplina que poseía los elementos necesarios para realizar investigación clínica con grupos de pacientes con el fin de evaluar el proceso diagnóstico y el pronóstico, así como para comparar los tratamientos. Asimismo, señaló que esta disciplina proveía en gran parte las bases necesarias para la construcción del edificio llamado “investigación clínica”; por esta razón, tituló su libro *La arquitectura de la investigación clínica* (1976).

Según Feinstein, el planteamiento de John R. Paul propició, en gran medida, el desarrollo de los métodos de la Epidemiología analítica y su extensión al campo de las enfermedades no transmisibles.

A través de su amplia obra escrita, Feinstein propuso los paradigmas de la investigación clínica, planteó que su principal objetivo era la predicción y no la explicación de los fenómenos; que la información obtenida de los pacientes era subjetiva en una gran proporción; que el reto no era el planteamiento de la hipótesis, sino la incorporación del mejor procedimiento para enriquecer el juicio clínico y que el método debía ser la observación ya que, por razones éticas, casi nunca es posible realizar experimentos.

Posteriormente, Sackett interpretó la Epidemiología Clínica como la disciplina en la que participa el médico dedicado al cuidado de los enfermos, provisto de información epidemiológica y estadística, para estudiar los procesos de diagnóstico, tratamiento y pronóstico; mencionó que la Epidemiología Clínica “es la aplicación de los métodos epidemiológicos y biométricos a la atención cotidiana del paciente”. Para cumplir con este propósito señalado por Sackett, el investigador clínico necesariamente debe utilizar diversas estrategias, de cuya aplicación adecuada dependa la validez de la información obtenida.

Por su parte, los Fletcher, en el decenio de 1980-1989, señalaron que para dar respuesta a las preguntas que surgen al estar frente a un paciente, es necesario recordar que la información clínica se basa en datos inciertos y que, por tanto, esto se expresa en términos de probabilidad, ya que si bien esta última se estima en relación con experiencias basadas en muestras de pacientes similares, las observaciones clínicas las efectúan médicos que usan instrumentos y técnicas diferentes y que, además, poseen sus propios juicios, con lo que se propicia la existencia de errores que alteran las observaciones. Surgió así la idea de que para contrarrestar las alteraciones de las observaciones clínicas, éstas deben apoyarse en principios científicos.

Como es evidente, el desarrollo vigoroso de la Epidemiología Clínica se inició a fines del decenio de 1970-1979; a partir de ese momento, con la irrupción de artículos y libros sobre el tema, se observó un creciente interés por esta disciplina.

La Facultad de Medicina de la UNAM, a principios del decenio de 1980-1989, dio un fuerte impulso a la Epidemiología; reorientó la maestría en ciencias médicas dentro de los cauces del método epidemiológico, incluyó la asignatura Epidemiología Clínica en el Plan de Estudios de 1985 de pregrado, la cual sigue vigente en el plan de estudios actual, y editó el libro de texto sobre esta disciplina. Fernando Cano Valle, el entonces Director de la Facultad, estaba convencido de la importancia de esta disciplina, así, nos conminó a un grupo

de médicos a diseñar dicha asignatura y a trabajar arduamente para la concreción de la que fue la primera edición del libro *Epidemiología Clínica*.

Esa primera edición salió a la luz en 1988, fue ampliamente aceptada no sólo por los estudiantes de Medicina, sino además por médicos clínicos, epidemiólogos y otros profesionales del área de la salud. La convicción de Cano Valle, médico neumólogo e investigador clínico, quien desde entonces reconocía ampliamente la importancia de la aplicación de la Epidemiología a la práctica e investigación clínica, señaló “[...] *Ciertamente, la tarea fundamental del médico es conocer las enfermedades del ser humano e investigar los medios para combatirlas. Reconocer los síntomas y hallar los signos mediante maniobras y razonamientos son principios en los que se ha sustentado la medicina, es decir, la clínica. Conforme se avanzó en el conocimiento de la enfermedad, la medicina se tornó en una diáspora de recursos para conocer y profundizar más en los problemas de salud del ser humano, de modo que la aplicación de ese conocimiento fuera puesta a la disposición de la sociedad. De esa manera, de la inspección, palpación, percusión y auscultación —reglas inmutables de la clínica y de la propedéutica—, así como de la observación de los enfermos y de la comparación y agrupamiento de las diversas manifestaciones de la enfermedad, nació la patología y en ese momento se transformó la medicina.*

*En esa transformación, tácitamente en relación con la investigación clínica, se aceptó de manera general que el médico es la persona responsable de planear la estrategia en el diagnóstico y de ejecutar las tácticas de la terapéutica; sin embargo, todavía hay quien duda de que cada aspecto del manejo clínico es diseñado, ejecutado y evaluado con procedimientos intelectuales idénticos a los que se emplearon en cualquier situación experimental.*

*Es evidente que para tratar un padecimiento, hay que reconocerlo. Saber qué se tiene que tratar, es decir, conocer el estado patológico del enfermo, en resumen, hacer el diagnóstico. Por ello el clínico recopila datos, de los cuales algunos le son de utilidad y otros no tanto, y en ocasiones la valoración e interpretación de éstos son parciales. Por lo que los elementos de juicio pueden ser incompletos. De ahí la importancia de la propedéutica (que en realidad es la base de la clínica), la cual origina la diferencia de hacer bien la práctica médica o no; pero la propedéutica ya no es suficiente en la actualidad. Si bien evita que el clínico se convierta en un autómata, ejecutando actos y acumulando datos, también enfatiza en el ser que piensa, agrupa datos, los sistematiza, interpreta, comprende y conforma en un diagnóstico. Sin embargo, la propedéutica carece de elementos que la evolución de la ciencia nos otorga. De ahí que del conocimiento de la metodología aplicada en los estudios que involucran al ser humano en el diseño y la interpretación validada por los métodos estadísticos dependan no sólo el éxito de una investigación, sino también la vida misma del paciente”.*

Asimismo, Cano Valle afirmó que “*la Epidemiología Clínica, a través de la aplicación correcta de la metodología epidemiológica, utiliza observaciones clínicas con las que puede establecer conclusiones válidas que respondan a las preguntas planteadas, salvando el obstáculo que representa la subjetividad, habitualmente presente en las mediciones realizadas en la práctica médica*”.

Es cierto que el término “Epidemiología Clínica”, su concepto y alcances causaron una importante polémica en su inicio; al margen de las polémicas suscitadas, es incuestionable que para la observación clínica, bajo el rigor científico, la Epidemiología Clínica constituye una de las mejores alternativas.

El interés que despertó la primera edición de este libro constituyó la motivación para elaborar la segunda edición en la que se actualizaron todos los temas, se incluyeron varios capítulos y se dio un enfoque más actual, más profundo y más humano.

En esa 2a. edición destacamos la importancia de la Epidemiología Clínica para el desarrollo de la investigación; asimismo, subrayamos que esta disciplina permite enfatizar que la ciencia, en su necesidad de esclarecer y generar conocimiento nuevo, se opone por principio a la opinión. Puede, en efecto, validarla planteándola a manera de suposición como lo es la hipótesis, es decir, como una respuesta tentativa a la pregunta de investigación. Sin embargo, si no se plantea correctamente dicha pregunta, no será posible concebir tal hipótesis o respuesta tentativa, por lo tanto no podrá generarse conocimiento científico.

Tuvieron que pasar varios años para que fuera posible sacar a la luz la tercera edición de este texto que, si bien mantiene la estructura de las ediciones anteriores, contiene varios capítulos nuevos, incorpora como autores a 13 destacados epidemiólogos y epidemiólogos clínicos y ofrece un panorama renovado de los alcances de la Epidemiología Clínica.

Al igual que los textos anteriores, se invita al lector a transitar por cada uno de los capítulos en los que encontrará los amplios horizontes teóricos, metodológicos e instrumentales que ofrece la Epidemiología Clínica.

Así, después de reseñar en el primer capítulo el desarrollo de la disciplina y su relación con la investigación, en el siguiente apartado se ofrece al lector algunos elementos para plantear preguntas de investigación a partir de delimitar el problema de estudio y, posteriormente, en un capítulo nuevo, se señalan los lineamientos para elaborar proyectos de investigación destacando de forma somera cada uno de sus apartados. Es importante recalcar esto último, ya que en toda investigación debe haber concordancia lógica entre los objetivos y el diseño metodológico utilizado, su análisis y la interpretación de los resultados. Ello implica además la selección adecuada de las técnicas estadísticas no sólo en las fases finales de la investigación (análisis e interpretación de resultados), sino también en la estructuración del diseño de la misma.

En un capítulo reformado se presentan los diseños de investigación, su conceptualización, algunos ejemplos de ellos y sus alcances y limitaciones. De la misma manera que en el libro anterior, un apartado especial fue destinado para los estudios experimentales, capítulo que fue renovado y actualizado ampliamente. Además, por la importancia que reviste en la investigación sobre fármacos, se incorporó el tema de Farmacovigilancia, cuyo propósito es orientar al lector sobre las medidas que se deben tomar para incrementar la seguridad para el paciente, y mostrar los métodos para la evaluación de los beneficios y los riesgos potenciales de cualquier intervención terapéutica.

Los capítulos de Bioética y Epidemiología Clínica y el de Revisiones sistemáticas y Metaanálisis en Medicina fueron modificados, este último enfatiza la importancia de proporcionar a los clínicos una respuesta objetiva para la toma de decisiones basada en el resumen de todas las evidencias disponibles.

El capítulo de Clinimetría en su nueva versión, destaca la importancia de todo el proceso de medición, la manera de controlarlo y evaluarlo con el fin de obtener información consistente, válida y confiable. Posteriormente, se presenta un apartado sobre la noción de Normalidad en Medicina, sus usos y formas de ser interpretada y calculada.

El capítulo Conceptos de causalidad en Medicina no sufrió grandes modificaciones debido a que su autora, la Dra. Florencia Vargas Voráckova, estaba atravesando por problemas de salud. Aprovecho este espacio para expresar mi reconocimiento a su labor profesional y cualidades personales, y mi tristeza por su partida.

Posteriormente se presenta el capítulo de Diagnóstico y evaluación de pruebas diagnósticas, en él se resaltan la importancia de reconocer la confiabilidad de las diferentes pruebas y la secuencia para establecer el diagnóstico de un padecimiento, además enfatiza la importancia de que los procedimientos diagnósticos deben sustentarse firmemente en la información obtenida a través del acto clínico.

El siguiente apartado hace referencia a la Estimación del pronóstico de la enfermedad, es decir, la evaluación del tiempo de ocurrencia de un evento relacionado a ella, además se destaca el análisis de la sobrevivencia. El capítulo de Análisis de decisión en la práctica médica brinda alternativas para contender con uno de los problemas más grandes que enfrenta el médico, la necesidad tomar decisiones médicas con base en información imperfecta o incompleta. Dos capítulos posteriores, Evaluación de la calidad de vida y de la atención médica, mediante el uso de la Epidemiología Clínica ofrecen un amplio panorama sobre la búsqueda de parámetros que permitan identificar y propone medidas para que la población viva con mayor bienestar y mejor atención médica.

Para cerrar el libro se incluyen tres capítulos; uno sobre el Plan de manejo estadístico, otro sobre el significado clínico y la significancia estadística y otro más sobre cálculo del tamaño de muestra.

Así, los autores de esta obra ofrecemos a la comunidad médica una tercera edición ampliada y actualizada que ha sido producto de la tarea cotidiana y de la experiencia diaria tanto en la investigación, como en la práctica clínica y en la docencia.

En este libro se pretende dar al estudiante de Medicina, de otras áreas de la salud, al de posgrado, al médico general o al clínico, los elementos metodológicos que le permitan realizar la práctica de manera más médica científica, más documentada y más sistematizada, sin olvidar nunca la gran importancia que tiene el componente humanístico en todo médico.

**Laura Moreno Altamirano**

# Prólogo a la segunda edición

## Avatares de la medicina occidental

¿Cuándo y cómo nació la medicina? Nada cuesta imaginar a un ser humano que, con curiosidad y compasión, se aparta de la ruta que sigue la tribu nómada y se acerca al desvalido que ha quedado atrás. Esta actitud inició un tipo superior de comercio humano: la relación médico-paciente que, muchos siglos después, el Dr. Ignacio Chávez, citando a Luis Portes, habría de sintetizar así: “una confianza frente a una conciencia”.

La medicina, o más bien, la concepción de la medicina — su marco conceptual — históricamente ha dependido de los contextos ideológico y político de la sociedad. No son de extrañar sus ligas con la religión y la magia cuando estas últimas bastaban para explicar al mundo. La palabra “medicina” contiene la raíz latina *med* que significa llevar al centro, equilibrar, balancear (palabras afines son “mediar”, “promedio”, “medida”); el término obliga entonces a ejercer una acción para cumplir con las funciones de equilibrio, de centrado. Este “equilibrio” se pierde con la enfermedad. El médico tiene la función de restablecer la salud y, por ello, los términos “medicina” y “terapéutica” son indisociables. De hecho, acudimos a la farmacia a comprar “la medicina” (no nos ofrecen a cambio la disciplina que estudia la enfermedad, sino un fármaco). Si la idea de la enfermedad es mágica o religiosa, la terapéutica se establecerá de acuerdo con lineamientos acordes con esa visión del mundo. Ese marco conceptual bien puede prescindir de la realidad.

En la antigua Grecia, donde todo nació, la escuela hipocrática diseñó “el concepto de los humores” con un alarde de armoniosa simetría. El fundamento de la vida, de acuerdo con esa concepción, depende de cuatro humores: sangre, flema, bilis amarilla y bilis negra. Estos humores no constituyen únicamente la base de la vida humana, sino que también explican que el temperamento y la salud dependen del equilibrio de estos “humores” en el cuerpo. La enfermedad es el exceso o el defecto en alguno de ellos y, por consecuencia, la terapéutica (la acción del médico, su razón de ser) debe orientarse al restablecimiento de la “armonía humoral”. La exposición de estas ideas es deliberadamente esquemática, porque lo que importa es resaltar sus consecuencias. Elaborada en el siglo VI a. de C., la teoría humoral de la enfermedad reinó única y soberana por más de 20 siglos. Es difícil calcular el número de muertos a consecuencia de sangrías, purgantes, lavativas, eméticos, ayunos prolongados y administración de tóxicos.

Ese estado de cosas empezó a cambiar durante el Renacimiento. Leonardo y Vesalio sistematizaron el conocimiento anatómico, y Antonio Benivieni escribió el primer libro conocido de correlaciones clinicopatológicas: *De Abditis Nonnullis ac Mirandis Morborum et Sanationum Causis*, y con ello inauguró una tradición seguida por Fernel, Boneto y muchos otros, que culminó con la publicación de la obra magistral de Morgagni: *De Sedibus et Causis Morborum per Anatomen Indagatis*. En sus vastas páginas se aprende que no hay tal trastorno en los “humores”; la enfermedad tiene una expresión en las lesiones de los órganos cuyo lenguaje se traduce en síntomas. Este lenguaje habría de ser descifrado hasta la segunda mitad del siglo XVIII, con el nacimiento de la clínica. Hubo entonces un sano repliegue de la terapéutica y se establecieron los fundamentos de la medicina moderna: la historia natural de la enfermedad, el diagnóstico diferencial y la correlación clinicopatológica.

## La ciencia más joven

Durante muchos años, el buen médico era el que diagnosticaba las enfermedades basándose en el conocimiento de los textos especializados y en su propia experiencia; proporcionaba además consuelo a los pacientes y familiares, siguiendo la tradición de la estirpe. En ocasiones, curaba; a veces, aliviaba. Se describe aquí a un hombre bueno, con sentido común, sabiduría linneana y conocimiento adquirido durante el envejecimiento. Pocas medidas terapéuticas de eficacia probada y la certeza de no saber el origen ni el remedio de los males que enfrentaba, matizaban su conducta ante el paciente. De acuerdo con los preceptos del juramento hipocrático, enseñaba su ciencia a otros y cultivaba su espíritu curioso con la descripción de nuevas

entidades nosológicas. Durante el siglo pasado y gran parte de éste, floreció la nomenclatura económica de la enfermedad (el riñón de Bright, la corea de Sydenham, la cirrosis de Laennec, el fenómeno de Lucio, el síndrome de Cushing, etc.) en catálogos nosográficos de cientos de páginas y como resultado, del esfuerzo para caracterizar procesos muy variados. Fue la época —de acuerdo con la terminología en boga— de los estudios descriptivos. La imagen descrita puede parecer idílica, pero no es irreal; al margen de críticas ácidas en torno al quehacer médico (véanse las obras de Montaigne, Moliere, Dickens y Shaw que se ocupan del tema), las expectativas del médico y de la sociedad a la que pertenecía pueden ilustrarse bien con el epitafio del Dr. Bright, que reza así:

*Sacred to the memory of sir Richard Bright, MD, DCL.  
Physician extraordinary to the Queen  
He contributed to medical science many scientific discoveries  
And works of great value  
And died while in the full practice of his profession  
After a life of warm affection  
Unsuited purity  
And great usefulness.<sup>1</sup>*

En nuestros tiempos, quizá ya no pedimos “cálidos afectos y pureza inmaculada” de nuestros médicos. Pero sí exigimos eficacia porque, en pleno siglo XX, la medicina deviene ciencia y lo hace en forma acelerada. La mutación ocurre cuando el vasto caudal de conocimientos médicos incorpora la información derivada de otras disciplinas biológicas que nacieron científicas: la microbiología, inmunología, genética, bioquímica, fisiología, anatomía patológica, farmacología, etcétera. Ocurre también cuando incorpora los avances tecnológicos al estudio de la enfermedad: la radiología, las técnicas de asepsia y antisepsia, el microscopio, la antibioticoterapia, la biología molecular, la informática, etcétera. De esa suerte, y aun cuando no haya generado todavía ninguna teoría, la medicina se convierte en la ciencia más joven.

Se postula en la actualidad que el médico debe realizar actividades asistenciales, de docencia y de investigación como parte de su código ético. De otra suerte, practica no una ciencia sino una técnica, un oficio. ¿Cómo cruzar el puente? ¿Cuál es la herramienta útil para trascender de la práctica diaria al conocimiento universal? Así como las artes aspiran a la precisión de la música, las ciencias aspiran a la precisión de las matemáticas. La ciencia parte de preguntas bien formuladas, de ahí a la definición y control de las variables, a la medición de los fenómenos y al conocimiento de su significado en un proceso de pensamiento que genera repetidamente el ciclo, a partir de una nueva pregunta. ¿Es posible usar el proceso científico en el fenómeno médico cotidiano? La respuesta es sí. El puente y la herramienta se llaman Epidemiología Clínica. Esta disciplina novedosa en el estudio de la enfermedad, ha permitido, una vez que se formula la pregunta de investigación, la sistematización coherente de los datos, el diseño cuyos resultados aproximen más el conocimiento a la realidad, su medición y su validación matemática. Hoy en día es difícil vivir sin los conceptos de normalidad, sensibilidad, especificidad, sesgo, variable, consistencia interna y tantos otros provenientes de la jerga epidemiológica. El conocimiento que ha generado la Epidemiología Clínica es ahora tan necesario como el de cualquier ciencia básica de la carrera del médico. Merece la pena detenerse en las metáforas “puente” y “herramienta”. La primera se usa aquí en el sentido de vía de acceso sobre un obstáculo; la segunda indica que, sin medicina, sin enfermedades, sin enfermos, la epidemiología clínica no tendría razón de ser; algo así como un taller de carpintería sin madera.

<sup>1</sup> Consagrado a la memoria de Sir Richard Bright, MD, DCL / Médico extraordinario de la Reina / Aportó a la medicina muchos descubrimientos científicos / Y trabajos de gran valor / Y murió en pleno ejercicio de su profesión / Después de una vida llena de cálidos afectos / Pureza inmaculada / Y de gran utilidad

## Un libro de Epidemiología Clínica

Si los párrafos anteriores son ciertos (y así conviene creerlo para la buena ejecución de este prólogo), entonces es bueno que aparezca una nueva edición del libro Epidemiología Clínica. Las convenciones del género prólogo, cuando éste es escrito por alguien que no es autor del texto, indican la necesidad de los énfasis sobre “llenar un vacío”, “obra esperada por mucho tiempo”, “de gran utilidad”, “en la vanguardia del conocimiento”, “dirigido a aprendices, oficiales y maestros” y otros lugares comunes por el estilo (salvo que los lugares comunes denuncien verdades evidentes).

El arduo placer que depara la lectura de los capítulos de este libro es evidencia de algunas obviedades: el texto en relación con su edición anterior se ha enriquecido en capítulos, temas y autores.

Los autores, en su mayoría jóvenes, escriben con conocimiento de causa y convicción. Hay numerosos ejemplos prácticos y una notable riqueza de referencias pertinentes.

Al margen de las obviedades, este esfuerzo coordinado por Laura Moreno Altamirano, Fernando Cano Valle y Horacio García Romero tiene un claro diseño: plantea dudas, muchas dudas; parte de la incertidumbre; no nos dice la verdad sino las aproximaciones a la verdad. Este diseño general es bueno porque el conocimiento nace de la duda y la incertidumbre. Para compensar y evitar la desesperación a la que también lleva la duda, en cada capítulo se ofrecen las estrategias las herramientas del tratamiento inicial de los problemas que plantea el estudio sistematizado de la enfermedad. Una virtud adicional: está escrito en castellano, que es nuestra lengua.

Esta *opus* de miembros distinguidos de la, aún no formal, Sociedad Mexicana de Epidemiología Clínica, se conocerá por sus frutos. Si en los tiempos por venir la influencia de esta obra propicia el incremento en la investigación clínica, el esfuerzo habrá merecido la pena. Es prudente terminar el prólogo con esa profecía.

**Óscar Larraza Hernández, 1994**  
**Jefe del Departamento de Anatomía Patológica,**  
**Hospital Central Sur de Alta Especialidad,**  
**PEMEX**

## Prólogo a la primera edición

Hace algunos años leí con detenimiento y gran interés en una de las mejores revistas de medicina, si no es que en la mejor, un artículo muy crítico sobre el futuro de la Epidemiología. El tema distaba mucho de corresponder a los tópicos que yo me veía obligado a leer cotidianamente para mantenerme al día en mis actividades y, sin embargo, el título lo hacía muy atractivo. Decía el autor que la Epidemiología estaba en peligro de desaparecer, no por falta de buenas, magníficas intenciones, sino por ser una ciencia en gran medida irrealizable. En pocas palabras, los augurios no eran buenos. Nunca olvidé tal artículo por varios motivos: primero, por lo bien escrito que estaba; segundo, porque se hacía una crítica seria a los límites de la ciencia en una de sus ramas, tema que luego adquiriría cierta popularidad y, finalmente, porque me parecía extraordinario que pudiera condenarse a la desaparición a una rama de la medicina con tan elegante facilidad.

El asunto no hubiera pasado a mayores y yo seguramente habría olvidado pronto tan notable escrito y sus conclusiones, de no ser porque mi actividad como inmunoinfectólogo y pediatra cada día me convenciría de lo contrario, es decir, del saludable vigor y de lo absolutamente esencial que estaba resultando la tan recientemente criticada Epidemiología. Si antes esta rama se abocaba al estudio de los episodios epidémicos de una enfermedad, como si esto fuera tan diferente de sus versiones endémicas o de su existencia interepidémica, hoy la Epidemiología estudia la enfermedad no como ocurre en un individuo, sino en cuanto a sus

peculiaridades de aparición en conjuntos de individuos, prestando atención a la distribución en el tiempo y el espacio, y las características de residencia, empleo, raza, sexo, edad, hábitos, etcétera, en los afectados. Como modesto partícipe en la emergencia de la sepsis neonatal por estreptococo beta hemolítico grupo B como causa fundamental de muerte por infección perinatal en los países desarrollados, pude verificar que sin un apropiado encuadre epidemiológico, todo lo demás que estudiábamos carecía de sentido, aunque también era cierto que la Epidemiología se nutría más que generosamente de los conocimientos de ramas tan vastas como la Microbiología, la Inmunología, etcétera. No tardarían las legionelas en causar sus estragos y en darnos un nuevo ejemplo de las virtudes de la Epidemiología a pesar de sus limitaciones. Sin deseos de bagatelizar, se antoja la Epidemiología como el detective en jefe de uno o más casos insolubles. Luego vino la gradual e inexorable emergencia del SIDA, enfermedad que ha tomado literalmente por sorpresa a la humanidad haciendo que afloren atavismos, metáforas y comportamientos ritualistas que ya creíamos relegados al pasado. Si resulta fascinante la inmunología del SIDA, más aún lo es su epidemiología, que no sólo nos da respuestas, sino que nos confronta con muchas preguntas que demandan atención urgente. El concepto global de esta enfermedad, su significado social y cultural, su influencia en el cambio de patrones de comportamiento, su carácter inexorable y el pánico que esto suscita nos hacen recordar actitudes que en el medio tuvieron nuestros antepasados frente a la peste. Con estos pensamientos de fondo es que respondo a mis colegas de la Facultad de Medicina que me han pedido escriba un prólogo a su extraordinario libro de Epidemiología, el cual marca un evento pionero en la medicina mexicana. Me defendí de tan inesperado e inmerecido honor, arguyendo que había personas más allegadas a la Epidemiología que podrían escribir un prólogo más relevante. Mi defensa fue inútil; sus razones fueron de más peso: realmente puede tener relevancia un comentario de alguien ajeno al campo, que no a sus implicaciones. Estoy seguro que este primer testimonio de lo que puede llamarse la Escuela Mexicana de Epidemiología, alojada en la Facultad de Medicina de la UNAM, tendrá un efecto seminal en las futuras generaciones de médicos mexicanos. La puerta principal que da acceso a una verdadera concepción social, justa y relevante de las enfermedades es, en mi opinión, el espíritu de este libro.

**Dr. Roberto Kretschmer**  
**Jefe de la División de Inmunología**  
**Unidad de Investigación Biomédica,**  
**Centro Médico Nacional “Siglo XXI”, IMSS**

## Agradecimientos

Al Ingeniero David Limón Cruz por su participación en la revisión minuciosa del libro.

A la Srita. Dolores Hernández González por su apoyo secretarial.

Al equipo de la Editorial McGraw-Hill por el excelente trabajo realizado; al Dr. Javier de León y al Lic. Emilio Salas; dedico una mención especial al Lic. Manuel Bernal Pérez.

Y en particular a los autores de cada capítulo y colaboradores de este libro por su confianza y paciencia, a los profesores que han impartido la asignatura de Epidemiología Clínica en la Facultad de Medicina de la UNAM, algunos de ellos por más de 20 años, por sus críticas y recomendaciones. Asimismo, a los alumnos que han trabajado las ediciones anteriores con este texto, por sus sugerencias y motivación para continuar.

**Laura Moreno Altamirano**





# Capítulo 1

## El desarrollo de la Epidemiología Clínica y su método

*Laura Moreno Altamirano*

*Sergio López Moreno*

La investigación científica es la principal actividad que el ser humano lleva a cabo para promover el avance del conocimiento y eliminar las prácticas y creencias mal fundadas. Su creciente capacidad para explicar el orden del mundo las ha colocado en un lugar semejante al que hace siglos tuvieron otros sistemas de pensamiento, como la religión y la mitología. Hoy se sabe que las enfermedades se deben a la ruptura de un complejo sistema de interacciones naturales y sociales, y que los padecimientos pueden limitarse e incluso eliminarse, siempre y cuando se cuente con las herramientas científicas y tecnológicas apropiadas. En algunos casos este desplazamiento de saberes ha sido un proceso más o menos lento, pero en general las ciencias forman el discurso dominante en materia de salud.

El predominio de la explicación científica frente a otras explicaciones sobre el mundo natural y social es actualmente casi absoluto, y se ha pasado de un mundo regido por la certeza religiosa a uno caracterizado por el predominio del racionalismo científico. Actualmente todos los procesos patológicos se explican científicamente, lo mismo que las modificaciones del ambiente o del mercado económico; la educación se basa en la descripción científica del mundo y un argumento vale tanto como “la cantidad de ciencia” que contenga. No es gratuito entonces que la cultura contemporánea gire en torno a las ciencias y que todos los medios políticos, económicos o culturales las aprecien tanto.

El camino que las ciencias utilizan para producir conocimiento se conoce como *método científico*, y todavía se discute si es sólo uno —que sería útil para todos los tipos de ciencia— o si realmente existe un método para cada ciencia o grupos de ciencias. Algunos filósofos de las ciencias sostienen que el método es sólo uno y que cada ciencia desarrolla variaciones de un único método, útil para todos los casos. Otros pensadores sostienen que cada forma de hacer ciencia implica un método científico particular, prácticamente inaplicable en otros campos científicos.<sup>1</sup> En favor de esta postura se argumenta que una de las formas de distinguir las diferentes ciencias es precisamente el tipo de método que

usan. Debido a que la manera de acercarse al estudio de cada objeto constituye, en pocas palabras, el método específico de una ciencia, puede hablarse de tantos métodos como ramas generales de la ciencia. Así, por ejemplo, aunque la Física y la Química estudien el mismo objeto, su diferente nivel de acercamiento determina su naturaleza específica y las coloca como ciencias diferentes. Esto mismo sucede incluso con saberes específicos dentro de cada ciencia (como en el caso de la Física teórica o de la Bioquímica). Es también frecuente poner fronteras a cada tipo de ciencia mediante diferenciar sus objetos de estudio y además completar esta distinción al señalar los modelos teóricos y principios que les son más o menos propios. No obstante, ninguno de estos criterios es un elemento suficiente para caracterizar y separar a una ciencia de otra.<sup>2</sup>

En resumen, pareciera que es relevante distinguir los diferentes campos de la ciencia mediante identificar el nivel de acercamiento y la perspectiva de la que parten, la particularidad de su objeto de estudio, los métodos que utilizan de manera preponderante y los modelos teóricos que han generado para ordenar sus sistemas conceptuales.

Esta breve incursión epistemológica parece necesaria para ubicar de manera más o menos precisa el sitio de la Epidemiología Clínica en el panorama general de las ciencias y, específicamente, en el campo particular de la investigación en salud.

## Investigación clínica

Durante los últimos dos siglos se consideró como investigación clínica a aquella que se realizaba directamente en sujetos humanos y que tenía como propósito conocer las características morfológicas y el curso fisiológico de la enfermedad a fin de aplicarlos en el diagnóstico, pronóstico y tratamiento de las personas enfermas a escala individual. En este caso, el nivel de acercamiento, la perspectiva elegida, el objeto de estudio y los métodos utilizados eran considerados típicamente clínicos; es decir, aplicados junto al lecho del enfermo (el *klinos* griego) y encaminados a generar conocimiento útil para modificar el curso clínico de la enfermedad individual.<sup>3-5</sup>

Aunque por supuesto que el conocimiento generado por la investigación clínica podía aplicarse a conjuntos de personas (sanas o enfermas), tanto sus métodos de acercamiento como su objeto de estudio (la enfermedad como proceso clínico individual) hacían que la investigación clínica tuviera una identidad propia y un alcance limitado. Durante más de 150 años los modelos que se usaron para ordenar esta forma de investigar fueron los que generaron la fisiopatología y la anatomía patológica, la observación clínica de signos y síntomas y la tecnología médica complementaria. Fue debido a la aplicación sistemática y generalizada de estos saberes como la práctica médica actual llegó a tener tanto éxito y pudo denominarse científica. No obstante, en la primera parte del siglo xx la investigación clínica comenzó a apoyarse en otros métodos igualmente específicos,

enriqueciéndolos conforme los adaptó a su particular objeto de estudio (la enfermedad como proceso clínico individual). Tal es el caso de los métodos de investigación generados por la Epidemiología, que surgió en el siglo XIX como un instrumento dirigido específicamente al estudio de la enfermedad en las poblaciones humanas; es decir, de la enfermedad considerada a escala colectiva.<sup>6</sup>

## La investigación epidemiológica

Desde su nacimiento formal, hace unos 150 años, se acepta que la Epidemiología tiene como propósitos explicar la dinámica de la salud de las poblaciones, así como las respuestas sociales adoptadas para conservarla o recuperarla. Para lograrlo describe y explica la ocurrencia y distribución de las condiciones de salud, busca descubrir su origen y propone medidas que buscan prevenir las enfermedades y sus consecuencias.

Desde el punto de vista práctico, la Epidemiología investiga la distribución y magnitud de los fenómenos patológicos, la forma en que varían entre las diferentes poblaciones y las circunstancias con las que se asocian. Ha sido definida de múltiples formas, pero todas las definiciones señalan que estudia los patrones de distribución de las enfermedades en las poblaciones humanas y los factores que influyen en ellos.<sup>7</sup> Las definiciones más recientes, como la de Kleinbaum, consideran que la Epidemiología describe el estado de salud de las poblaciones e identifica su magnitud; identifica la frecuencia y tendencias de la enfermedad entre diferentes grupos; explica la etiología, factores asociados y modos de transmisión de las enfermedades; predice el curso de la enfermedad en las poblaciones y propone medios de control usando medidas preventivas y de erradicación.<sup>8</sup> James Frost, pionero de la medición epidemiológica, la consideraba una ciencia inductiva interesada no sólo en describir la distribución de la enfermedad, sino en generar una filosofía paralela sobre la salud en general. Como quiera que sea, la Epidemiología es una disciplina integradora, ecléctica, que para estudiar la enfermedad en grupos humanos aprovecha los conceptos y métodos provenientes de otras ciencias como la Biología, la Estadística, la Sociología, la Psicología o la Economía, entre muchas otras.

Desde el punto de vista teórico, los principales aportes de la Epidemiología se refieren a los modelos explicativos del curso de las enfermedades (historia natural de la enfermedad); a los modelos explicativos y predictivos de los procesos infecciosos (especialmente durante las epidemias); a los modelos de transición epidemiológica, y a los modelos que explican el comportamiento social de los procesos morbosos relacionados con la sobrevivencia humana (como la teoría de la comprensión de la enfermedad).

Es claro que, por lo menos en apariencia, la investigación clínica es distinta de la investigación epidemiológica, por lo menos sus objetos de estudio, métodos y principales teorías. ¿Cómo es entonces que la investigación clínica y la investigación epidemiológica se combinaron de tal forma que terminaron desarrollando una nueva disciplina, la Epidemiología Clínica?

Aunque a lo largo de los siglos XVIII, XIX y XX la Medicina Clínica y la Epidemiología aparecieron como disciplinas diferentes, originalmente se desarrollaron de manera muy

estrecha. Los fundadores de la Epidemiología fueron médicos clínicos. No fue sino hasta el siglo XVIII cuando ambas ramas del saber médico se establecieron como dos áreas distintas, con intereses específicos. De acuerdo con Kenneth Rothman, los estudios epidemiológicos a gran escala, iniciados en Estados Unidos en el siglo XX, tuvieron desde su inicio profundas repercusiones en la práctica clínica.<sup>9</sup> En 1914, Joseph Goldberger desarrolló un estudio clásico sobre la pelagra que demostró que esta enfermedad no era una infección, y que su tratamiento clínico más apropiado debía ser la ingesta de proteínas animales. Desde ese momento la investigación epidemiológica se amplió a tal grado que dejó de considerarse una disciplina restringida al estudio de brotes epidémicos y enfermedades infecciosas. En 1920, comenzó en Massachusetts el estudio epidemiológico de las enfermedades crónicas, y el mismo año Inglaterra empezó un estudio sobre la epidemiología del cáncer. En el decenio de 1920-1929, Broderso y Lane Claypon estudiaron factores asociados con el cáncer epidermoide y el cáncer de mama, respectivamente.

En el decenio de 1930-1939, Dean llevó a cabo un estudio sobre caries y fluoridación del agua, con el que comenzó la era de las grandes investigaciones de campo. En 1935 Greenwood publicó un libro clásico que incluía el estudio epidemiológico de la tuberculosis y otras infecciones, pero también del cáncer y las enfermedades mentales. En 1938, John R. Paul utilizó por primera vez el término “epidemiología clínica”; Alvan Feinstein afirma que ese año dio inicio el desarrollo del método epidemiológico aplicado a la clínica médica.<sup>10</sup>

Más tarde se llevaron a cabo diferentes estudios que conformaron las bases de la Epidemiología Clínica. Destacan el estudio Framingham sobre enfermedades cardiovasculares, iniciado en 1949; el ensayo de campo de la vacuna Salk, en 1954; el de Wynder y Leven de 1950 sobre consumo de tabaco y cáncer pulmonar,<sup>11</sup> y los de Richard Doll y Bradford Hill, quienes ese mismo año empezaron una serie de trabajos observacionales que demostraron 25 años después que la relación tabaco-cáncer era una relación de causa-efecto.<sup>12</sup> Varios estudios que arrojaron importantes aportaciones para el estudio de enfermedades crónicas siguieron desarrollándose en las décadas de 1960-1969, 1970-1979 y 1980-1989, mientras que la Epidemiología Clínica continuaba su vigoroso desarrollo (cuadro 1-1).

Alrededor de 1960 el término “**Epidemiología Clínica**” se aceptó académicamente como una disciplina con legitimidad propia.

En los últimos decenios la clínica y la epidemiología han seguido combinándose, lo que ha dado como resultado el cuerpo teórico-metodológico específico de la Epidemiología Clínica. Esta conjunción ha permitido que la práctica clínica cuente con más elementos para otorgar su justa dimensión a las observaciones clínicas, seleccionar las mejores pruebas diagnósticas e interpretarlas de manera correcta, elegir la secuencia lógica en la estrategia diagnóstica, juzgar en forma objetiva los resultados de la terapéutica elegida, emitir pronósticos con bases más sólidas, comprender mejor lo que se lee en las publicaciones médicas periódicas y transferir los resultados de la literatura a la atención de pacientes.<sup>13</sup>

**Cuadro 1-1.** Desarrollo histórico de la Epidemiología Clínica

|           |                          |                                                                       |
|-----------|--------------------------|-----------------------------------------------------------------------|
| 1914      | Goldberger, J            | Pelagra y factores socioeconómicos                                    |
| 1920      | Broderso                 | Cáncer epidermoide y hábito de fumar pipa                             |
| 1926      | Lane Claypon, JE         | Cáncer de mama y factores asociados                                   |
| 1938      | Dean, HT                 | Fluorosis y caries dental                                             |
| 1935      | Greenwood                | Tuberculosis y otras infecciones, cáncer y enfermedades mentales      |
| 1947      | Schreck, R y Lenowitz, H | Cáncer de pene, circuncisión y hábitos higiénicos                     |
| 1947      | Sartwell, P.             | Hepatitis B y transfusión sanguínea                                   |
| 1948      | Kennaway, EL             | Cáncer de útero y factores sociales                                   |
| 1950      | Gagnon, F                | Etiología del cáncer uterino                                          |
| 1950      | Sheridan, MD             | Rubéola en el embarazo                                                |
| 1950      | Wynder, EL               | Cáncer broncogénico y tabaquismo                                      |
| 1950      | Diversos                 | Leucemia, cáncer de mama, vejiga, cuello uterino, pulmones y estómago |
| 1951      | Doll, R y Hill, AB       | Mortalidad y hábito de fumar                                          |
| 1954      | Salk, JE                 | Vacuna contra la polio                                                |
| 1955      | Doll, R                  | Cáncer de pulmón y asbesto                                            |
| 1949/1959 | Dawber, TR               | Estudio Framingham (enfermedad cardiovascular)                        |
| 1956/1960 | Morris, JN               | Cardiopatía isquémica en conductores de autobús                       |
| 1956      | Last, JM                 | Fluoridación del agua y caries                                        |
| 1962      | Mc Carrol, JR            | Accidentes automovilísticos fatales                                   |
| 1966      | Bizzozero, OJ            | Leucemia y radiación en Hiroshima y Nagasaki                          |
| 1968      | Speizer, FE y Doll, R    | Mortalidad por asma                                                   |
| 1970      | McMahon, B               | Cáncer de mama y edad al primer parto                                 |
| 1978      | Kelsey, J                | Tabaquismo materno y malformaciones congénitas                        |
| 1979      | Rooks, JB                | Uso de anticonceptivos orales y adenoma hepatocelular                 |
| 1980      | Linos, A                 | Radiación y leucemia                                                  |

A este cuerpo de conocimientos se le ha denominado también como **clínica bioestadística**, **razonamiento clínico** y **clinimetría**, entre otros nombres. En este libro se considera a todos ellos como sinónimos.

En su desarrollo, la Epidemiología Clínica ha debido enfrentar algunas dificultades. Una de las primeras fue el rechazo inicial de los investigadores clínicos para aceptar la realización de investigaciones bajo los postulados del método epidemiológico. Aunque cada vez es más raro, aún ocurre que se conciba a la Epidemiología como un campo administrativo cuyo único objetivo es el desarrollo de acciones de salud pública y de administración

de servicios médicos. No obstante, con la enorme irrupción de textos y artículos de Epidemiología Clínica, este problema ha disminuido. En toda esta literatura se sigue aceptando que la investigación médica puede realizarse bajo las modalidades de investigación biomédica, clínica o sociomédica y que, realizadas de manera adecuada, todas ellas son igualmente científicas. Se trata de modelos de producción de conocimiento que resultan complementarios.

## La Epidemiología Clínica

Los conceptos fundamentales de la Epidemiología Clínica derivan de los aportes de la epidemiología clásica, con la diferencia de que la epidemiología clínica se realiza al lado del enfermo, junto a la cama del paciente. La mayoría de los estudios en este campo tiene como propósito fundamental mejorar el ejercicio del clínico.

Al respecto, Jenicek señala que la Epidemiología Clínica permite mejorar las decisiones clínicas a la cabecera del enfermo, organizar y estructurar la investigación clínica y conocer la lógica y la arquitectura de los estudios que se publican en las revistas médicas; permite detectar las grandes fallas, valorar la fuerza de la evidencia presentada y decidir si la información es apropiada para su aplicación práctica a los pacientes.<sup>11</sup>

El método de la Epidemiología Clínica, como afirma Rothman,<sup>14</sup> posee bases teóricas y técnicas coherentes, y se ha ganado un sitio como disciplina científica. Este método consiste en general en comparar la probabilidad de que ocurra un evento (casi siempre patológico) en poblaciones que poseen por lo menos una característica diferente entre sí (por lo regular, la causa presumible del fenómeno patológico). En otras palabras, lo que compara la Epidemiología Clínica es la diferente probabilidad de ocurrencia de enfermedad en poblaciones que difieren entre sí por alguna característica, presumiblemente la causa de las diferencias de probabilidad. La observación clínica complementa el arsenal metodológico de la Epidemiología Clínica. Como puede notarse, la base de la inferencia epidemiológica se encuentra en el concepto de relación causal. Su método, como el de todas las ciencias, tiene como sustento la proposición y comprobación de hipótesis causales. La estrategia tradicional de la Epidemiología es la inferencia inductiva, que consiste en generalizar para la población estudiada los fenómenos ocurridos en una muestra de ella.

Con el desarrollo de las críticas al método inductivo (según las cuales ninguna cantidad de experiencia puede ser usada para fundamentar relaciones sólidas de causa-efecto), las ciencias empíricas debieron recurrir a una estrategia común: la probabilidad. El uso de modelos probabilísticos es entonces una aplicación indispensable en el método epidemiológico empírico. Puede decirse que el intento final de la Epidemiología Clínica es medir de la manera más precisa el efecto del “azar” en la presentación y desarrollo de los fenómenos patológicos que investiga. Al ser imposible esta medición en términos absolutos, lo que hace es calcular exactamente la probabilidad de que sea el azar el que causa los fenómenos observados; así, elimina a este último como componente causal válido de enfermedad. Al no aceptar los sucesos aleatorios como componentes de los mecanismos causales de la enfermedad, el método de epidemiología se ha convertido, con ayuda de

las técnicas estadísticas, en la mejor manera de promover el avance del conocimiento médico clínico, al colocar al azar como componente causal que se ignora o no ha sido identificado. Esta ignorancia puede corregirse “conforme el conocimiento se expande”, según las palabras de Rothman.

La Epidemiología Clínica es la aplicación de los principios y método epidemiológico a los problemas concernientes a la Medicina Clínica con el fin de brindar una mejor atención, realizar investigación y comprender críticamente la literatura médica.

Existe un amplio campo de la Epidemiología Clínica que no trata específicamente del desarrollo y la evolución de las enfermedades. La validación de la eficacia de diferentes técnicas diagnósticas y medidas terapéuticas, los estudios de calidad de vida y calidad de la atención, el análisis de decisiones clínicas y los estudios de metaanálisis han recibido una atención cada vez más intensa. Estas actividades de investigación son impensables en la actualidad sin la participación de la Epidemiología Clínica. En su desarrollo participan también los conceptos de causa-efecto, probabilidad de ocurrencia y efecto del azar. Naturalmente, esta investigación hace uso también de las técnicas de medición y procesamiento estadístico de la información. Empero, sus resultados tienen un efecto más directo sobre la práctica clínica y pasan, casi de inmediato, a ser utilizados por el profesional de la medicina en su actividad cotidiana.

El conocimiento de la enfermedad a través del estudio del sujeto enfermo es tarea que pertenece de manera exclusiva al médico clínico. La observación aguda y prudente, capaz de sistematizar signos y síntomas en síndromes y enfermedades; la elección de la mejor medida terapéutica para cada paciente; la elaboración de un pronóstico hipotético y la creación de un vínculo afectivo con el enfermo que permita crear una confianza capaz de influir en el curso de su enfermedad, no pueden realizarse fuera del seno de la medicina clínica.

El clínico ocupa entonces un lugar insustituible en la investigación que realiza la Epidemiología Clínica, y es muy posible que ello nunca deje de ser así. En 1969, cuando era más que evidente el predominio de los físicos, químicos y biólogos en la recepción de los premios Nobel de Medicina y Fisiología, el doctor Ignacio Chávez se preguntaba si la predominancia de la investigación biomédica en el campo de la salud no estaría anunciando el ocaso de la clínica, y él mismo respondía señalando: “[...] seguramente que no. Mientras la Medicina sea Medicina humana, no habrá ocaso de la clínica. Es ella la que sirve de base a los nuevos estudios y es la cúspide a la que convergen los nuevos avances”. Esta respuesta, casi medio siglo después, se mantiene vigente.

Como es evidente, la aplicación del método epidemiológico a la práctica clínica ha mostrado sus innumerables virtudes. Todo indica que los avances de esta disciplina serán cada día mayores y que sus principales logros están apenas por realizarse. Si con ello el arte del diagnóstico, el pronóstico y el tratamiento —y, por ende, el futuro de los pacientes— resultan beneficiados, la empresa vale la pena.



## Referencias

1. Geymonat, L. *Límites actuales de la Filosofía de la Ciencia*. Gedisa, Madrid, España. 1987.
2. Almeida Filho, N. *La Ciencia Tímida. Ensayos de deconstrucción de la epidemiología*. Lugar Editorial/Universidad Nacional de Lanús. Buenos Aires, Argentina. 2000.
3. Pérez TR. Qué es la investigación clínica y dónde se ubica dentro de la investigación científica. En: Alarcón SD *et al.* *Fundamentos de la investigación clínica*. Siglo XXI, México, pp. 9-20, 1988.
4. Fletcher RH, Fletcher SW. Clinical epidemiology: a new discipline for an old art. *Ann Intern Med* 99:401-403. 1983.
5. Lilienfeld AM, Lilienfeld D. *Foundations of epidemiology*. 2a ed., Oxford University Press, Nueva York, EUA, 1979.
6. Buck C, Llopis A *et al.* *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. OPS, Washington, 1988.
7. Susser M. *Causal thinking in the health sciences: concepts and strategies of epidemiology*. Oxford University Press, Nueva York, EUA, 1973.
8. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiology, principles and quantitative methods*. Life Time Learning Publications, Belmont, EUA, 1982.
9. Rothman JK. *Modern epidemiology*. Little Brown, Boston, EUA, 1986.
10. Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. WB Saunders, Filadelfia, EUA, 1985.
11. Jenicek M. *Epidemiología. La lógica de la medicina moderna*. Masson, Barcelona, España. 1996.
12. Doll R, Hill AB. Smoking and carcinoma of the lung. *Br Med J* 2:739-748. 1950.
13. Wachter N, Lifshitz A. Qué es la epidemiología clínica y para qué le sirve al clínico. *Rev Médica IMSS Méx* 27:171-174. 1989.
14. Wynder EL, Graham LA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *J Am Med Assoc* 143:329-338. 1950.

## Bibliografía

- Bizzozero OJ, Johnson KG. *Radiation-related leukemia in Hiroshima and Nagasaki 1946-1964. Distribution, incident and appearance time*. New England Journal of Medicine 1966;274 (20):1095-1101.
- Broders AC. *Squamous-cell epithelioma of the lip*. Journal of the American Medical Association 1920;74:656-664.
- Cornfield J. *A method of estimating comparative rates from clinical data*. Journal of the National Cancer Institute 1951;11:1269-1275.
- Dawber TR, Kannel WB, Gordon TO. *Coffee and cardiovascular disease: observations from the Framingham Study*. New England Journal of Medicine 1974;291:871-874.
- Doll R, Hill AB. *Smoking and carcinoma of the lung*. British Medical Journal 1950;2:739-748.
- Doll R, Hill AB. *A study of the etiology of carcinoma of the lung*. British Medical Journal 1952;2:1271-1286.

- Goldberger J. En: Buck C, Llopis A *et al.* *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. Washington: OPS, 1980;630-658.
- Lane-Clayton JE. A further report on cancer of the breast. En: Schlesselman JJ. *Case control studies: design, conduct, analysis*. New York: Oxford University Press, 1982.
- Levin ML, Goldstein II. *Cancer and tobacco smoking. Preliminary report*. Journal of the American Medical Association 1950;143:336-338.
- Linos A, Gay JE, Orves AL. *Cow-dosis radiation and leukemia*. New England Journal of Medicine 1980;302:1101-1105.
- Sheridan MD. En: Buck C, Llopis A *et al.* *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. Washington: OPS, 1980;723-731.
- Schreck R, Lenowitz H. *Etiology factors in carcinoma of the penis*. Cancer Research 1947;7:180-187.
- Wynder EL, Graham EA. *Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma*. Journal of the American Medical Association 1950;143:329-338.

# Capítulo 2

## La fundamentación del problema de investigación

*Horacio García Romero*

*Pablo Kuri Morales*

*Saúl León Hernández*



### Introducción

El punto de inicio de una investigación es la identificación de la pregunta que se quiere contestar. Habitualmente, cuando se presentan los resultados de una investigación, muy pocas veces se dice al lector cómo nació en el investigador la idea de realizarla.

Pero si se pidiera al investigador, cualquiera que sea su campo científico, que describa lo que hizo durante su actividad de investigación, él contestaría casi de manera invariable que ha “respondido a una pregunta”.

Podría decirse que detrás de toda investigación hay siempre una pregunta y que lo que el científico hace es intentar contestarla. Richards sintetiza lo anterior definiendo al investigador como “un ser humano que camina con una pregunta bajo el brazo”.

La pregunta constituye, entonces, no sólo la guía que permite diseñar metodológicamente el camino de la investigación; es, en pocas palabras, la parte medular de la investigación. Todo acto de creación científica, por tanto, implica la resolución de un problema. Pero, ¿cómo surge un problema científico?, ¿existe un programa de reglas que conduzcan al planteamiento de problemas científicos?, ¿este programa es lineal, lógico y ordenado?

### El problema de investigación

El científico y filósofo Peter Medawar, premio Nobel de Medicina, responde que no a la pregunta planteada en el párrafo anterior. No existe un sistema único y uniforme que indique la manera en que puedan “descubrirse” y plantearse problemas científicos. Quizá esta sea la razón por la que, al revisar un texto sobre Metodología, se halle un esquema que detalla un orden riguroso, lineal, lógico y hasta acartonado de cómo efectuar una investigación, pero que pocas veces se mencione el proceso a partir del cual las preguntas sobre la realidad cotidiana se convierten en científicas.

Para abordar las condiciones en las que una pregunta se convierte en un problema científico, es necesario, sin negar la importancia de la formalización, enfatizar el carácter creativo (a veces lógico, intuitivo y azaroso) de la actividad científica. ¿De qué depende la creación científica?

Dice Bunge que, en primer lugar, el investigador requiere estar inmerso en el campo de conocimientos que conforman el marco en el que el problema se presenta. Es decir, es necesario que pueda “moverse” con facilidad entre la información que existe sobre el tema. Un médico jamás se preguntará sobre un problema matemático complejo, a menos que en la práctica sea también un matemático, y viceversa. Además, cuando un profesional se hace una pregunta sobre un tema extraño a su campo de actividad, es muy posible que tal problema esté ya resuelto y que constituya no una laguna del conocimiento en general, sino una laguna en el conocimiento del propio profesional. A veces, al investigador le toma toda la vida encontrar una pregunta de investigación que valga la pena ser contestada. La exploración, la reflexión y la búsqueda parecen ser el destino del científico.

## La pregunta de investigación

De lo anterior resulta que para que una pregunta sea científica, su respuesta debe ser desconocida para toda la comunidad científica. Debe cubrir lo que se llama una “laguna objetiva del conocimiento”. Por el contrario, las lagunas “subjetivas” del conocimiento (las que son propias del sujeto), se resuelven con la lectura de un buen texto o artículo.

No obstante, difícilmente planteará problemas relevantes quien desconozca a fondo el área en la que se aplica ni sepa los principios, teorías y conceptos de su campo profesional. Tampoco propondrá explicaciones alternativas ante el surgimiento de un hecho inesperado o el hallazgo de incongruencias entre las diversas teorías establecidas.

Sin embargo, saber mucho no basta, pues no garantiza el encuentro de un problema científico, y mucho menos su planteamiento adecuado. Para ello, dice de nuevo Medawar, “el investigador debe ser capaz de dudar de manera sistemática de lo que se da por cierto y de desconfiar de lo establecido”. En ocasiones, según Bunge, de poco sirve saber mucho si el saber no alcanza para plantear algo nuevo. El científico, el verdadero, por lo general es una mezcla de erudito, coleccionista, detective y aventurero. Se acerca con una gran compulsión a los límites del conocimiento aceptado sólo para romperlos después (o intentar romperlos).

Desafía, a veces de manera obstinada, el saber tradicional proponiendo nuevas formas de abordar la realidad. En este último sentido, hacer ciencia implica cierta dosis de audacia. Lo anterior no significa, empero, que el científico sea un hombre extraordinario. Las características anteriores valen para su actividad profesional. Fuera de ésta, el hombre de ciencia es tan común como cualquier ser humano.

Quizá una característica que sí es especial en el investigador es su disposición al trabajo. La investigación significa a veces mucho tiempo invertido, que parece desperdiciado cuando la respuesta no llega o llega mal. El trabajo científico en ocasiones no sólo es arduo y penoso, sino infructuoso.

En suma, hacer investigación requiere experiencia, dominio teórico del campo de trabajo y sobre todo una fuerte dosis de creatividad, paciencia y autocrítica.

Si es evidente que todo proceso de investigación implica la existencia de un hueco en el conocimiento objetivo o una contradicción en las explicaciones sobre el mundo, ¿por qué razón no todos pueden plantearse adecuadamente una pregunta científica?

Todo parece indicar que el planteamiento de problemas, además de ser un ejercicio creativo de búsqueda, observación, imaginación, síntesis e invención inmersa en una realidad parcialmente desconocida, se ajusta a ciertos modelos que los propios científicos han elaborado sistematizando su experiencia en la búsqueda de aproximaciones más finas y mejores a la verdad. Estos modelos resultan de la experiencia acumulada en decenas de generaciones de científicos y forman parte, aunque pocas veces se mencione, de la estrategia inventada por los investigadores para acercarse más fácilmente a sus objetos de estudio. Esta estrategia se conoce genéricamente como **método científico**.

Si, como Ackoff afirma, la mitad de la investigación consiste en el planteamiento adecuado del problema, la otra mitad comprende la aplicación correcta del método. Un problema planteado con claridad y un método desarrollado de manera correcta constituyen la mejor garantía de la calidad de una investigación.

Sin embargo, no siempre es posible plantear claramente un problema. Ello se debe a que con frecuencia el científico sólo tiene una idea general y vaga sobre el mismo. Lo anterior se resuelve cuando el investigador revisa la información existente sobre el tema de estudio, delimitando de modo cada vez más fino las relaciones que existen entre su aún confuso problema de investigación y el conocimiento aceptado sobre el mismo.

Conforme el investigador se acerca a los límites objetivos del conocimiento, le resultará cada vez más claro su problema de investigación. Con ello, es posible que para ese momento sepa realmente qué es lo que intenta hacer, cumpliendo un principio fundamental en el planteamiento de problemas que enuncia Kerlinger de la siguiente manera: “si se desea resolver un problema, es preciso saber de qué problema se trata”.

## La hipótesis y el problema de investigación

A medida que el investigador posee mayor información sobre el problema que intenta delimitar, necesariamente se acerca a las diferentes explicaciones ya existentes sobre los fenómenos involucrados. Si no existe una explicación satisfactoria sobre el fenómeno que investiga, o hay varias y son contradictorias, el investigador debe elaborar una explicación propia. En pocas palabras, es necesario que el investigador defina una hipótesis.

Las hipótesis son explicaciones tentativas, provisionales sobre un determinado fenómeno. Se caracterizan porque expresan las relaciones probables que hay entre dos o más variables. En la investigación científica constituyen un elemento central, ya que sin ellas es imposible guiar un proceso de investigación. Sin explicaciones tentativas a su problema, una investigación es, sencillamente, impensable.

Las hipótesis facilitan el tratamiento científico inicial de un problema debido a que además de permitir establecer relaciones entre variables, indican claramente la manera

en que estas relaciones pueden ser verificadas. Ambos elementos expresan de manera implícita qué variables deben investigarse y cómo pueden, potencialmente, ser medidas.

Así que las hipótesis científicas son por naturaleza comprobables a fin de demostrar su verdad o falsedad, constituyen un complemento del problema de investigación y están profundamente relacionadas con éste y con el marco teórico que hay a su alrededor.

Dewey afirma que la importancia del problema es similar a la de las hipótesis en cuanto a sus posibles respuestas. Señala que si la investigación comienza con una situación problemática, que deja inicialmente perplejo al científico, el problema sólo puede ser enunciado cuando se ofrece una explicación tentativa a esta situación problemática.

Por otra parte, las hipótesis ayudan a delimitar el problema en la medida en que formulan relaciones reducibles a dimensiones prácticas. Un problema muy general o vago será inmediatamente detectado cuando no sea posible comprobar la propuesta hipotética a través de la medición de variables operacionales. Por lo regular, cuanto más específico es un problema, más fácil es expresar conjeturas sobre su solución y disminuir las variables a cifras prácticas.

Finalmente, las hipótesis poseen una característica muy importante para no prescindir de ellas al plantear un problema: tienen la capacidad de predecir fenómenos. Cuando el científico afirma que “si ocurre X, sucederá Y”, lo que hace es una predicción causal. Si logra que cuando ocurra X, ocurra también Y, confirmará la hipótesis.

## El problema científico en Medicina

El trabajo cotidiano del médico clínico lo expone de manera continua a hechos y manifestaciones que difícilmente puede explicar o que sólo explica en forma parcial. Esto permite un continuo ejercicio de indagación y análisis de la información, sobre todo cuando el médico es consciente de que cada paciente puede aportar información que lleve a cuestionamientos nuevos, cuya resolución propicie conocimientos útiles.

En Medicina, el planteamiento de un problema puede basarse en diversos intereses, por ejemplo, establecer las magnitudes de una variable biológica, conocer la utilidad de un nuevo método diagnóstico o de tratamiento, pronosticar de manera más certera el curso de un padecimiento o la probabilidad de que se presente una complicación.

Sin embargo, también es posible que el médico sólo intuya la presencia de relaciones entre varios signos o síntomas que se describen aislados, o bien que quiera indagar el valor clínico de un dato inesperado de laboratorio. A veces, muy pocas por cierto, al médico no le satisfacen las explicaciones que hay para un fenómeno patológico, o bien encuentra incongruencias entre varios hechos o razones “científicamente” establecidos.

El planteamiento de problemas médicos, entonces, surge de observaciones, intuiciones y razonamientos, pero siempre deriva de la observación de situaciones que la teoría no contiene, que no espera o que contradice.

Kretschmer señala que “[...] el investigador primeramente reconoce los hechos y, por un proceso de selección (de lo que considera relevante) y supresión (de lo que considera irrelevante) genera una pregunta coherente. Es probable que aquí se encuentre la parte más genuinamente creativa de la investigación científica”.

## Requisitos de un problema de investigación

Una vez que el investigador ha logrado delimitar con más claridad el problema, su marco conceptual y las hipótesis probables sobre su respuesta, el planteamiento debe reunir varios requisitos para garantizar el logro de la investigación. Aunque varios de ellos han sido mencionados, en el cuadro 2-1 se listan todos porque resumen de manera adecuada las propuestas de la mayoría de los autores.

### Cuadro 2-1. Requisitos que debe reunir el planteamiento del investigador

- **Es recomendable presentar el problema en forma de pregunta.** A veces esto no es posible, pero siempre que lo sea debe llevarse a cabo porque es la manera más sencilla de saber que el problema está suficientemente delimitado.
- **El planteamiento debe señalar de manera clara cada uno de sus términos.** Si el planteamiento es confuso, será difícil precisar los objetivos, elegir el diseño y analizar la información recopilada.
- **El problema debe ser relevante y pertinente.** Es necesario que la respuesta al problema planteado resuelva alguna duda importante sobre el tema que se estudia. El costo en tiempo, dinero y esfuerzo invertidos para resolverlo debe justificarse por la importancia de las conclusiones que se espera obtener. El investigador que no toma en cuenta este concepto y que se embarca en proyectos de investigación banal e intrascendente, malgasta su vida y derrocha recursos que en otras manos podrían producir reales beneficios.
- **El problema debe ser específico.** El investigador ha de concentrarse en un número limitado y pequeño de variables o procesos de un fenómeno. La dispersión impide los registros adecuados, multiplica los errores y disminuye la confiabilidad de los resultados. Al plantear un problema, el investigador debe reducirlo a sus aspectos fundamentales, de manera que pueda garantizar su estudio exhaustivo.
- **La resolución del problema debe ser factible.** Todo problema ha de llevar implícita la posibilidad de su resolución. Esto significa que se cuenta con los medios metodológicos y los recursos tecnológicos, humanos y financieros, el tiempo y el espacio para intentar resolverlo. Si no se poseen estos requisitos, el problema debe desecharse como motivo de estudio.
- **El problema debe ser congruente con los conocimientos que la ciencia ha determinado como cercanos a la verdad.** Queda fuera del espíritu científico plantearse problemas que no estén basados en un mínimo de conocimientos o principios establecidos. Sin embargo, en áreas como la Biología y, por tanto, en la Medicina, es necesario reconocer que aún se está lejos de entender muchos de los fenómenos que son objeto de su estudio y que es posible, en ocasiones, plantear problemas con fundamentos que pueden parecer endeble.
- **El problema debe resolverse con base en la cuantificación de una o más variables en estudio, o bien, por medio de la interpretación de algún proceso.** En el primer caso, la respuesta del problema debe apoyarse en el análisis o la comparación de magnitudes. Aun aquellos factores que puedan parecer sólo cualitativos, son susceptibles de medición en escalas ordinales o semicuantitativas. De ahí que en el problema deba quedar implícita la pregunta: ¿qué elementos deben medirse? Si no hay factores que cuantificar, la pregunta toma un cariz cualitativo y el problema científico requiere de metodologías apropiadas.
- **Cada uno de los conceptos del problema debe definirse en términos operacionales.** Esto implica que ha de ser posible establecer las características de cada elemento del problema mediante procedimientos confiables y repetibles, al alcance de los sentidos. O bien, si se trata de un estudio cualitativo, deben esclarecerse las categorías de análisis.



Si bien este apartado se refiere básicamente a la investigación llamada cuantitativa, vale la pena recordar que, según varios autores, en la actualidad no se puede seguir hablando de un solo método, como se apunta en el capítulo 1. Así lo señala Moreno Altamirano.

En la actualidad se reconoce, sin lugar a dudas, que las ciencias de la salud cubren un amplio espacio: desde acontecimientos de dimensiones moleculares hasta el conglomerado social. Se evidencia cada vez más el origen de las enfermedades y se conoce con mayor profundidad el funcionamiento del ser humano mediante el estudio de las bases celulares y moleculares. También es cierto que la salud y la enfermedad cobran su sentido e incluso su fisonomía con la participación de procesos sociales, y que la comprensión cabal del ser humano requiere de la integración de conocimientos de muy distintos orígenes.

Así, cuando el análisis cuantitativo con las más innovadoras técnicas estadísticas parecía haber llegado a la cúspide, se hace evidente que ha dejado sin resolver problemas cruciales en la comprensión de las dimensiones biológicas del individuo en la dinámica histórica de los procesos sociales.

Consecuentemente, investigadores destacados en el enfoque cuantitativo comenzaron a promover los métodos cualitativos. Desde hace aproximadamente una década su uso y aceptación han sido inusitados, en gran parte, debido a los límites que se han advertido en la comprensión de los fenómenos derivados de prácticas, técnicas y protocolos cuantitativos, que han provocado los infortunados resultados del enfoque cuantitativo en la comprensión de los procesos de surgimiento, diseminación o preservación de perturbaciones significativas de la salud en ámbitos sociales determinados.

En los últimos decenios se ha ido gestando un cambio silencioso en la metodología de las ciencias humanas y sociales. Un renovado interés y una necesidad sentida por la metodología cualitativa han ido apareciendo entre sociólogos, educadores, antropólogos, psicólogos y médicos, entre otros. El enfoque cualitativo objeta la exigencia, en gran parte insostenible, de cuantificar toda realidad humana. Se ha hecho patente la frecuente irrelevancia y equívocos de la cuantificación y de la incompreensión cabal de innumerables procesos de salud producto de ignorar la importancia que tienen el significado, el contexto y la finalidad de las acciones humanas.

Por lo anterior es posible identificar dos grandes posturas en los enfoques actuales para el estudio de las ciencias humanas; por una parte, aquellas ciencias que persisten en el uso del método cuantitativo con fundamentos nítidamente empiristas, objetivantes, orientadas por métodos nomotéticos y en búsqueda de leyes generales, con criterios definidos en marcos de paradigmas estrictamente disciplinarios para aproximarse a los fenómenos, y por la otra, quienes buscan aproximaciones comprensivas al fenómeno humano, a sus múltiples determinaciones, a la complejidad de sus procesos, a la variedad de las culturas y los universos simbólicos para, en esa tentativa, derivar de las situaciones específicas el método adecuado.

*Laura Moreno Altamirano*

Los siguientes son ejemplos de problemas bien planteados acordes al enfoque cuantitativo: ¿qué grado de utilidad tiene el praziquantel en el tratamiento de la cisticercosis cerebral?, ¿qué datos de laboratorio pueden orientar cuando un cuadro de colangitis

ascendente se considera de pronóstico malo?, ¿cuáles son las cifras normales de la  $\text{PaO}_2$  en la arteria radial del recién nacido sano?, ¿cuáles son las percepciones sobre la gravedad de la enfermedad de un paciente con diabetes?

Algunos ejemplos de problemas planteados de manera deficiente serían los siguientes:

- ¿Cómo funciona el instinto de agresión? Este problema incluye términos difíciles de definir y una infinidad de variables a considerar. La respuesta a esta pregunta requiere del planteamiento de cuestionamientos más simples y de la participación de metodología cualitativa.
- ¿Se puede prevenir la aparición del cáncer? Un problema expresado así no tiene especificidad, es vago y carece de las bases para darle una respuesta.
- ¿La vitamina  $\text{B}_{12}$  es mejor que la vitamina  $\text{B}_1$  en el tratamiento de la onicomiosis? La incongruencia con los conocimientos ya establecidos resta todo valor a este problema. Además, operativamente sería muy difícil definir el término “mejor”, que es un concepto subjetivo que depende de un juicio de valor.

Rojas Soriano recomienda que al plantear un problema:

- Se conceptualice con precisión y se señalen sus límites teóricos. Las teorías que puedan explicarlo deben estudiarse a fondo, vinculando el planteamiento del problema con el cuerpo teórico en el que se moverá la investigación.
- Se fijen los límites temporales y espaciales de la investigación. Además, es necesario determinar el lapso requerido para analizar el problema planteado, el lugar geográfico en el que se intentará resolver y si se dará o no seguimiento al mismo.
- Se sitúe en los contextos tecnológico, económico e histórico en que se encuentra inmerso.

De acuerdo con Kerlinger, el adecuado planteamiento de un problema debe formular la pregunta de investigación con claridad y sin ambigüedades; expresar la relación entre dos o más variables, y realizarse de tal forma que pueda comprobarse empíricamente.

El siguiente ejemplo muestra cómo, a partir de la observación de un fenómeno no esperado por la teoría y con un planteamiento adecuado del problema, puede llenarse una laguna del conocimiento.

## Ejemplo del origen de un problema

En 1980 un grupo de médicos de Atlanta, en Estados Unidos, observó que el *Morbidity and Mortality Weekly Report*, publicado por un centro de enfermedades infecciosas, informó de un aumento en el uso de pentamidina, medicamento antiparasitario poco utilizado tradicionalmente.

La investigación indicó que cinco jóvenes sufrían de una enfermedad poco frecuente y rara: neumocistosis, tratada a base de pentamidina. Ante tal situación, el grupo se planteó la siguiente interrogante: “¿Por qué y en qué condiciones está aumentando la frecuencia de neumocistosis, lo cual también produce incremento en el uso de pentamidina?”

Ante tal interrogante, los investigadores iniciaron una revisión de la literatura y encontraron que se habían comunicado cuatro epidemias por ese padecimiento. La primera, en niños de Varsovia con alto grado de desnutrición; la segunda, en un grupo de niños prematuros; la tercera, en enfermos de cáncer tratados con fármacos inmunosupresores, y la cuarta, en un grupo de enfermos con trasplantes de órganos, también bajo tratamiento con inmunosupresores.

Los investigadores buscaron las características comunes en estos cuatro grupos.

El agente etiológico de la neumocistosis es *Pneumocystis carinii*, un parásito que infecta a muchos seres humanos sin causarles ningún problema, excepto cuando hay una respuesta inmunológica disminuida (factor común en los cuatro grupos comunicados en la literatura). Sólo en estos casos, el parásito se multiplica y produce los síntomas propios del padecimiento. Por esta razón, la neumocistosis no se considera una enfermedad contagiosa.

Con estos elementos los investigadores llegaron a la siguiente conclusión: si la neumocistosis no es contagiosa y ésta sólo se ha incrementado en pacientes inmunodeprimidos, entonces la inmunosupresión es la causa del aumento de la parasitosis. Ahora, el problema ha comenzado a ser planteado con un enfoque científico.

Los investigadores en seguida se preguntaron: ¿qué factores no evidentes han aumentado a partir de 1980 para que el número de pacientes inmunodeprimidos se haya incrementado? Postularon primero que la causa de la inmunodepresión, que a su vez producía la neumocistosis, podía ser una forma de cáncer, un linfoma o una leucemia. Sin embargo, la observación descartó esta primera hipótesis, y se puso entonces atención a las características epidemiológicas del padecimiento: las entrevistas y exámenes médicos informaron que todos los pacientes eran homosexuales, que sufrían de candidiasis, estaban invadidos por citomegalovirus y no presentaban causa evidente de inmunosupresión. Más tarde, algunos sujetos desarrollaron un tipo de cáncer poco común: sarcoma de Kaposi.

La neumocistosis siguió aumentando durante 1981, cuando se comunicaron 54 casos. Además, había ya 47 personas con sarcoma de Kaposi, siete con ambas enfermedades y un total de 108 casos con padecimientos antes poco comunes como infecciones por citomegalovirus, herpes y candidiasis. La mayoría de los enfermos eran homosexuales, del sexo masculino y de entre 25 y 50 años de edad.

En 1982 se encontró el síndrome en usuarios de drogas intravenosas. Se descartó al citomegalovirus como causa del problema, ya que se conocía desde hacía tiempo el cuadro que este germen originaba.

Los estudios clínicos y epidemiológicos orientaban a que la enfermedad era infecciosa y que se transmitía de manera predominante por vía sexual, y con alta probabilidad de ser causada por un virus.

Durante 1983 se realizaron infinidad de estudios al respecto. Ya nadie dudaba de que la cuestión planteada entre 1980 y 1982 por los médicos del Centro de Enfermedades Infecciosas de Atlanta fuera, en efecto, un problema científico. El descubrimiento del agente causal de la entidad detectada era sólo cuestión de tiempo.

Pocos años después, el científico francés Luc Montaigner aisló el virus llamado más tarde virus de la inmunodeficiencia humana (HIV), causante del síndrome de inmunodeficiencia adquirida (SIDA), una de las más recientes enfermedades descubiertas por el ser humano.

En poco más de 10 años de su aparición se sabía más de esta entidad que lo que ningún científico, hacía unos cuantos años, habría imaginado que sería posible conocer en tan poco tiempo sobre un solo padecimiento. Sobre el SIDA se conoce en la actualidad prácticamente lo mismo que lo que se sabe de enfermedades que tienen siglos o milenios asolando a la humanidad.

De alguna manera, el ejemplo de lo que se ha realizado alrededor del SIDA es un buen modelo de lo que puede hacerse en cualquier otra área de la ciencia. El ejemplo puede multiplicarse por 100 o 1 000. Toda la historia del pensamiento científico consiste en hacerse una buena pregunta y responderla de manera correcta, y aunque nadie puede asegurar que los conocimientos actuales serán igualmente válidos dentro de los años que siguen (de hecho, lo más probable es que suceda lo contrario), no hay duda de que por el momento constituyen la manera más adecuada de interpretar la realidad.

---

## Bibliografía

- Ackoff R. *Methods of inquiry*. Educational Publishers, St. Louis Missouri, EUA, 1950.
- Bunge M. *Teoría y realidad*. Ariel, Barcelona, España, 1971.
- De Gortari E. *El método de las ciencias*. Grijalbo, México, 1979.
- Dewey J. *Logic: the theory inquiry*. Holt, Rinehart and Winston, NY, EUA, 1938.
- Garza MA. *Manual de técnicas de investigación para estudiantes de ciencias sociales*. El Colegio de México, México, 1988.
- Kerlinger FN. *Behavioral research*. Holt, Rinehart and Winston, NY, EUA, 1973.
- Kretschmer R. Las preguntas y cómo contestarlas. En: Alarcón SD et al. *Fundamentos de la investigación clínica*. Siglo XXI, México, 1988.
- Medawar PB. *Los límites de la ciencia*. Fondo de Cultura Económica (Breviarios), México, 1988.
- Nachmias D, Nachmias C. *Research methods in the social sciences*. St. Martin Press, NY, EUA, 1987.
- Richards S. *Filosofía y sociología de la ciencia*. Siglo XXI, México, 1987.
- Rojas SR. *Guía para realizar investigaciones sociales*. Plaza y Valdés, México, 1987.
- Selltiz C. *Métodos de investigación en las relaciones sociales*. Rialp, Madrid, España, 1968.
- Vázquez CL. *El método científico en la investigación en ciencias de la salud*. Méndez Oteo, México, 1987.



## Capítulo 3

# Lineamientos para el diseño del proyecto de investigación

Laura Moreno Altamirano

### Introducción

El proyecto o protocolo de investigación es el documento en el que se describe la planeación de las diferentes etapas de una investigación.

Habitualmente se le ha llamado a este documento **protocolo**, del griego *protokollon*, la primera hoja de un papiro que describe los procedimientos para llevar a cabo alguna actividad.

Dicho término es utilizado en los países de habla inglesa, aunque con otra connotación. La Real Academia Española lo define en una acepción como “Plan escrito y detallado de un experimento científico, un ensayo clínico o una actuación médica”; de hecho, el uso de este vocablo está tan ampliamente difundido que en este escrito se utilizarán ambos términos (protocolo y proyecto) de forma indistinta.

La investigación científica es una de las actividades más elaboradas del pensamiento humano; en el capítulo 1 se señaló que es la principal actividad realizada por el ser humano para promover el avance del conocimiento y eliminar prácticas o creencias mal fundadas.

Entendemos por conocimiento el saber consciente y fundamentado que somos capaces de comunicar y discutir, el cual corresponde al término griego *episteme*, que se distingue del conocimiento vulgar o *doxa*.

El conocimiento científico es aquel obtenido mediante una metodología y de acuerdo a algún método científico. Este conjunto de conocimientos que se tiene sobre el mundo, así como la actividad humana destinada a conseguirlos, es a lo que se llama **ciencia** (del latín *scire* que significa: “saber, conocer”, y del griego *sophia*, “el arte de saber”).

La ciencia no debe perseguir la ilusoria meta de que sus respuestas sean definitivas, ni siquiera probables. Su avance se encamina hacia una finalidad infinita: la de descubrir incesantemente problemas nuevos, más profundos y justificar nuestras respuestas al respecto. Como afirma Bunge, “La ciencia no pretende ser verdadera ni, por tanto, final, incorregible y cierta”.

Con este fin, la investigación debe ser debidamente planeada, para lo cual es indispensable la elaboración del proyecto o protocolo.

El protocolo de investigación permite al investigador explicar el razonamiento científico; ordenar sus ideas en relación con el problema de estudio; determinar a través de la definición de los objetivos e hipótesis lo que se pretende obtener con la investigación; establecer el método, las técnicas y procedimientos más adecuados, así como definir las características de los sujetos en estudio y garantizar su seguridad. Asimismo, es útil para establecer los recursos necesarios y el tiempo requerido para la ejecución de cada una de las etapas y así evitar tomar decisiones arbitrarias en el desarrollo de la investigación.

## Componentes del protocolo de investigación

Es de suma importancia el hecho de que la *Ley General de Salud* en materia de Investigación para la salud, en su artículo 102 reglamenta la obligatoriedad de elaborar el proyecto de investigación, documento que sirve para que los Comités de Investigación y Ética de las diferentes instituciones evalúen y dictaminen la pertinencia, factibilidad y seguridad de la investigación que se pretende realizar (véase capítulo 7).

Si bien existen diferentes formatos, por lo general deben incluir los elementos descritos en el cuadro 3-1.

### I. Fundamentación

#### Título

La función del título es identificar el objeto de estudio. Se recomienda usar el menor número de palabras posibles, pero que describan adecuadamente el contenido. La brevedad no debe interferir con la claridad; cuando el título es muy largo lo mejor es redactar uno corto y complementarlo con un subtítulo.

Es conveniente iniciar el título con una palabra clave, de preferencia con un nominativo que describa o indique directamente el objeto de estudio, y completarlo con otros nombres, adjetivos o verbos, para así formar una frase breve, clara, concisa y que, a la vez, sea informativa. El título de ser posible debe contestar las preguntas: “¿qué?”, “¿a quién?”, “¿cuándo?” y “¿dónde?”

#### Antecedentes

Los antecedentes constituyen el marco de referencia que permite conocer el estado actual del problema planteado. Brindan orientación teórica, la cual permite construir el marco teórico o de referencia o modelo o esquema conceptual; es decir, el conjunto de conceptos relacionados que representan la naturaleza de la realidad.

La revisión de la literatura tiene por objeto que el investigador se actualice y profundice en el conocimiento existente respecto al problema planteado, conozca lo que otros

**Cuadro 3-1.** Componentes del protocolo de investigación

|                                                                   |
|-------------------------------------------------------------------|
| <b>I. Fundamentación</b>                                          |
| • Título                                                          |
| • Antecedentes                                                    |
| • Planteamiento del problema                                      |
| • Pregunta de investigación                                       |
| • Justificación                                                   |
| • Marco teórico                                                   |
| • Hipótesis                                                       |
| • Objetivos                                                       |
| <b>II. Metodología</b>                                            |
| • Diseño del estudio                                              |
| • Población o muestra                                             |
| ▪ Características y número                                        |
| ▪ Métodos de selección de la muestra                              |
| ▪ Variables y escalas                                             |
| • Procedimientos de recolección de datos                          |
| ▪ Métodos                                                         |
| ▪ Fuentes                                                         |
| ▪ Técnicas                                                        |
| ▪ Instrumentos                                                    |
| ▪ Formas de medir                                                 |
| • Organización y análisis de datos                                |
| ▪ Recuento y presentación tabular y gráfica                       |
| ▪ Medidas de resumen (tendencia central y dispersión)             |
| ▪ Pruebas estadísticas                                            |
| • Lineamientos éticos                                             |
| ▪ Carta de consentimiento                                         |
| <b>III. Implementación</b>                                        |
| • Recursos y logística                                            |
| ▪ Recursos humanos, metodológicos, materiales, técnicos y físicos |
| ▪ Cronograma de actividades                                       |
| <b>IV. Referencias bibliográficas</b>                             |

autores han investigado al respecto, los métodos utilizados y los resultados obtenidos, con la finalidad de ahorrar horas de trabajo, de sugerir nuevas hipótesis o bien mejorar el diseño de las investigaciones previas, y de evitar cometer los mismos errores o repetir estudios innecesariamente.



**Cuadro 3-2.** Fuentes de datos que debe cubrir la bibliografía

- 
- Bases electrónicas. MEDLINE (Pubmed) reúne varias bases de datos de literatura científica como el Index Medicus.
- 
- Otras bases de datos: Excerpta Medica, Current Contents, Science Citation Index, Biblioteca Cochrane, LILACS, etcétera.
- 
- Artículos de investigación original, de revisión, de opinión o crítica y libros en relación con el problema en estudio.
- 
- Registros y reportes especiales como: estadísticas vitales y censos.
- 
- Reportes de grupos de trabajo especialistas en el problema de estudio, como las series de reportes técnicos y consultas con expertos en el área.
- 

Es preciso que los antecedentes incluyan una revisión sistemática (actualizada y relevante) de la literatura existente sobre el problema de investigación planteado, por lo que deberán contener resultados o hallazgos de estudios preliminares, nacionales y extranjeros (véase capítulo 8).

La bibliografía debe ser estudiada de manera cuidadosa para que las referencias sean críticas y selectivas; el cuadro 3-2 muestra algunas características que debe contemplar.

Las referencias deben presentarse acotadas de acuerdo con el Index Medicus. Cada concepto o enunciado tiene que referir la cita bibliográfica correspondiente; ésta puede indicarse con un número encerrado entre paréntesis o mencionar el apellido del autor o autores. La referencia deberá escribirse de manera completa en la sección de referencias bibliográficas.

## Planteamiento del problema y pregunta de investigación

El planteamiento del problema de estudio es la descripción clara de lo que se propone conocer, probar o resolver mediante la investigación. El problema debe reflejar la relación entre la observación empírica y el sustento teórico, por lo que la lectura crítica de la literatura es fundamental. Del planteamiento del problema deben desprenderse lógicamente los objetivos y las hipótesis.

Dicho de otra manera, del problema surge la “pregunta” que da origen al estudio. La palabra “problema” y sus equivalentes en otros idiomas designan una dificultad que no puede resolverse automáticamente, esto es, con la sola intervención de los reflejos condicionados o aprendidos.

Los problemas en investigación son preguntas que se plantean en las esferas del hacer (problemas prácticos) y del conocer (problemas teóricos).

Un problema puede consistir, ya sea en llenar un vacío existente en el conocimiento, o en resolver una contradicción entre dos o más propuestas.

Así, es necesario encontrar, crear o construir algún objeto material o ideal, o bien demostrar, refutar, confirmar o rechazar una proposición o un sistema de proposiciones.

Advertir problemas que los demás pasan por alto, plantearlos con claridad, insertarlos de manera adecuada en un cuerpo de conocimiento y resolverlos con el máximo rigor posible con el propósito primordial de enriquecer el saber, son los cometidos del investigador científico “problematizador” por excelencia.

Un problema puede considerarse planteado científicamente si se formula en términos inteligibles y precisos, si se dispone de un fondo de conocimientos en el que se pueda eslabonar y si puede abordarse mediante algún procedimiento (método o técnica) disponible o creable.

Decía Emmanuel Kant: “lo que es dato para la sensibilidad, es tarea para el intelecto”. El entendimiento no se conforma con los datos que le proporcionan los sentidos: los elabora, los pone en duda y crea nuevos objetos. El término “dato” es algo desafortunado, pues sugiere una actitud pasiva. No todo dato nos es dado; la experiencia cotidiana provee datos propiamente dichos; pero la experiencia científica busca nuevos datos y hasta los produce deliberadamente, como ocurre en la experimentación. Además, en la esfera del conocimiento no hay dato sin problema ni problema sin dato. En efecto, la investigación científica es una actividad consistente en advertir, plantear y resolver problemas y ninguno de estos tres momentos se da en el vacío, sino que supone un conjunto de datos, algunos de los cuales figuran de manera explícita en los enunciados de los problemas.

Es imposible plantear un problema cuando no se sabe nada; mientras que cuanto más se sabe, mayor es el número de problemas que se puedan plantear. En la ciencia, de poco sirve saber mucho si el saber no alcanza para plantear problemas nuevos.

Rara vez uno se plantea problemas para cuya solución carece de métodos adecuados; por otra parte, si no se cuenta con dichos métodos surge un nuevo problema: el de diseñar nuevos métodos.

La naturaleza de los problemas científicos no está determinada sólo por la naturaleza de los objetos a que se refieren. La propia selección de los problemas está determinada por los conocimientos disponibles, por intereses teóricos (confirmación o refutación de una teoría), por posibilidades metodológicas (disponibilidad de una técnica adecuada) y aun por supuestos filosóficos.

Los problemas no “surgen”, “se plantean” ni “se dan” de manera impersonal. Son los individuos, con sus conocimientos y prejuicios, quienes los formulan.

A partir del problema se debe, en la medida de las posibilidades, plantear la pregunta de investigación, misma que es la guía que permite elegir el método y diseñar metodológicamente el camino de la investigación (véase capítulo 2).

La pregunta es **“la médula de la investigación”**.

Para plantear preguntas relevantes es necesario estar inmerso en el campo del conocimiento del problema, así como conocer los principios, teorías y conceptos de ese campo profesional.

## Justificación

Constituye las razones por las que se ha decidido llevar a cabo la investigación; en esa sección deben señalarse los beneficios que se esperan. Se especificarán la relevancia y trascendencia del estudio en el ámbito de lo social, sus implicaciones prácticas para contribuir a resolver un problema y su aportación teórica para llenar un hueco en el conocimiento.

## Marco teórico

Corresponde a aquel en el cual se ubica la investigación. Es un marco conceptual conformado por conceptos o conjuntos de conocimientos que interrelacionados dan lugar a teorías y respuestas hipotéticas relacionadas en forma sistemática y coherente para dar una explicación o predecir algunos fenómenos.

Las teorías se van conformando a través de diferentes niveles de desarrollo; la descripción, el análisis y la explicación. Así, el sustento teórico del estudio implica analizar y proponer teorías de acuerdo con algún enfoque teórico.

El rigor científico se manifiesta en la coherencia lógica de todo el proceso de investigación y de la estrategia teórica utilizada, es decir, de bases teóricas y conceptuales, lo cual permite generar un conocimiento válido.

El marco teórico es el sustento del problema de investigación y ayuda a precisar y organizar los elementos contenidos en la descripción del problema (cuadro 3-3).

## Hipótesis

Es la respuesta tentativa al problema planteado y es indispensable que esté sujeta a comprobación. La hipótesis científica debe desprenderse como consecuencia de la experiencia personal del investigador y por deducción de resultados ya conocidos. La relación entre el cuerpo de teoría y los datos empíricos da lugar al planteamiento de la hipótesis. Es un enunciado factible de ser contrastado.

Además, a partir de la hipótesis es factible plantear sus consecuencias verificables, lo que ayudará a identificar las variables, el diseño y los métodos estadísticos necesarios

### Cuadro 3-3. Funciones del marco teórico

- 
- Ayuda a delimitar el área de investigación, es decir, seleccionar hechos conectados entre sí, mediante una teoría que dé respuesta al problema formulado.
  - Sugiere guías de investigación. En la elaboración del marco teórico pueden verse nuevas alternativas de enfoque para tratar el problema; puede incluso cuestionarlo.
  - Expresa proposiciones teóricas generales, postulados, marcos de referencia, mismos que van a servir como base para formular hipótesis, operacionalizar variables y procedimientos y técnicas a seguir.
  - Amplía el horizonte del estudio al compendiar el conocimiento existente.
  - Ayuda a prevenir errores u omisiones que se han cometido en otros estudios.
  - Provee un marco de referencia para interpretar los resultados del estudio.
-

para su comprobación, es decir, propone conjeturas o supuestos provisionales acerca de la relación entre dos o más fenómenos o variables.

La hipótesis prevé las bases para la investigación y al plantearla es preciso especificar ampliamente las condiciones de comprobación o contraste (véase capítulo 2).

En conclusión, la hipótesis es una explicación tentativa sobre la relación entre variables. A partir de ella, el contrastarla se plantea en esta forma: “Si esto ocurre, entonces los resultados serán ...”.

La hipótesis debe ser congruente con el planteamiento del problema en cuanto:

- Definición operacional de los términos que la conforman.
- Sustento en conocimientos comprobados.
- Armonía con los conceptos científicos sobre el tema.
- Debe incluir una explicación sencilla pero suficiente sobre los hechos.
- Debe conducir de manera racional a la predicción teórica de los hechos.
- Las consecuencias previstas deben ser verificables.

## Objetivos

Representan la finalidad que persigue la investigación; es decir, los logros directos y evaluables que se pretende alcanzar. La definición de objetivos claros, precisos, pertinentes, factibles y trascendentes constituye un paso esencial posterior al planteamiento del problema de investigación. Para definir los objetivos se deben tomar en cuenta los aspectos mencionados en el cuadro 3-4.

Los objetivos deben estar directamente relacionados con las consecuencias de la hipótesis. Es importante tener especial cuidado en la selección de los verbos que describen la acción que se pretende lograr (tales como “conocer”, “evaluar”, “comparar”, “determinar”, etc.) y, de ser posible, deben ser jerarquizados.

Los objetivos deben ser:

- Claros.
- Pertinentes.
- Factibles.
- Trascendentes.
- Precisos.
- Congruentes entre sí.

**Cuadro 3-4.** Factores indispensables para definir los objetivos

- 
- El interés específico de la investigación.
  - Los recursos físicos, humanos y financieros.
  - La metodología adecuada para comprobar la hipótesis.
  - Todas las posibles actividades adicionales que participan en el proceso de investigación.
- 

## II. Metodología

El término “metodología” designa cómo se enfocan los problemas de investigación y se busca darles respuestas, es decir, señala la manera de realizar la investigación.

A través de la metodología se define cómo se **aprehende, ordena, interpreta y analiza** la realidad estudiada. La postura filosófica y epistemológica acerca de la ciencia, de la que parte el investigador, orienta la elección metodológica.

La metodología se entiende como la parte del proceso de investigación que permite sistematizar los métodos y las técnicas necesarios para llevarla a cabo. La metodología da pie a la selección de métodos y técnicas concretos de investigación. Los métodos son vías que facilitan la obtención de conocimientos para comprender un problema de investigación.

Así, indica el “cómo” se llevará a cabo la investigación, con el fin de llegar a cumplir los objetivos y a la demostración o verificación de la hipótesis, es decir, constituye una descripción detallada de los métodos y técnicas que se utilizan durante la investigación. Incluye los siguientes aspectos:

### Diseño del estudio

Una vez definida la sección de fundamentación, será posible seleccionar el diseño más adecuado para conducir la investigación. Para definir el tipo de estudio a realizar se debe tomar en cuenta lo siguiente: la pregunta que se quiere contestar, los objetivos y las hipótesis planteadas.

En un inicio se debe definir si se trata de un estudio **experimental** u **observacional**.

Cuando es un estudio observacional, es necesario identificar si es **comparativo** o **descriptivo; transversal** o **longitudinal**, o bien, **retrospectivo** o **prospectivo**.

Los estudios experimentales se caracterizan por ser **comparativos, longitudinales** y **prospectivos**.

Es importante revisar los criterios para la clasificación de estudios en el capítulo 4, así como las características de los estudios **transversales, casos y controles, cohorte** y **experimentales**. A partir de esta revisión cuidadosa se podrá elegir correctamente el diseño más apropiado.

### Población o muestra

La población objetivo o universo es definida como el grupo de individuos o unidades de observación a los que se extrapolan los resultados obtenidos en la investigación.

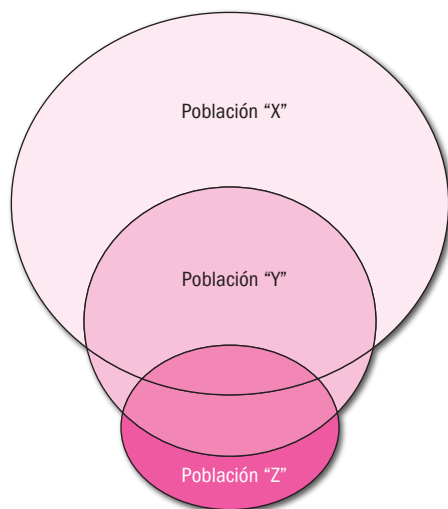


**Figura 3-1.** Ejemplo de gráfica para determinar una población muestra.

Lo fundamental en la definición anterior es que para determinar la población se requiere especificar la serie de características comunes que deben poseer los individuos, o unidades que la conforman. Cuando las características comunes a los elementos de la población son pocas, se tendrá una población más amplia; por el contrario, cuando las características comunes son muchas, la población se reduce. De este modo, el concepto de la población es flexible y depende de las características comunes de los elementos que la conforman.

En el ejemplo de la figura 3-1, la población X es parte de la población Y, y la población Y es parte de la población Z. Se dice que la población Y tiene mayor grado de generalidad que la X y menor que la Z. Como regla general, al tener mayor grado de generalidad (menos características comunes en los elementos) la población tiene más variabilidad, esto es, los elementos tienen mayor diferencia y las mediciones efectuadas en ellos tendrán fluctuaciones mayores. Esto se debe a que mientras más características comunes tengan los elementos de una población, más parecidos serán.

## Muestra

La muestra es un subconjunto representativo de la población, que se utiliza cuando no es factible estudiar a toda la población. En la mayoría de las investigaciones es imposible estudiar a toda la población pues ello implicaría, entre otras cosas, una elevación exorbitante de los costos y de los tiempos para su realización. Así que por lo general, en los trabajos de investigación cuantitativa, no así en los estudios cualitativos, el interés radica en extrapolar los resultados de la muestra a la población general; cuando es posible hacerlo se dice que el estudio tiene validez externa (véase capítulo 19). El cuadro 3-5 muestra cuándo es posible realizar la extrapolación.

**Cuadro 3-5.** Factores indispensables para poder extrapolar a la población objetivo

- 
- Que la muestra sea representativa de la población; para ello debe ser adecuada en características y en número, además de haber sido seleccionada de manera aleatoria.
  - Que la muestra difiera de la población en sus características esenciales.
  - Que las características en las que difieren no sean relevantes para el estudio.
- 

A fin de definir la población y la muestra se deben especificar los criterios de inclusión, exclusión y eliminación.

- **Criterios de inclusión.** Características que deben estar presentes en las unidades de estudio para que sean consideradas como parte de la población y puedan ser incluidas en el estudio.
- **Criterios de exclusión.** Son los que, de estar presentes en las unidades de estudio, hacen que no pertenezcan a la población y, por ello, evitarían su inclusión.
- **Criterios de eliminación.** Características que, al presentarse durante el estudio, obligarían a eliminar a esa unidad de la investigación.

Para definir el número adecuado de unidades a estudiar es necesario calcular el tamaño de la muestra (ver capítulo 19).

**Métodos de selección de la muestra.** En este inciso debe definirse el marco de muestreo para cada población que se quiera estudiar. El marco de muestreo constituye el sitio de referencia donde se podrá localizar a todas las unidades de la población. El marco de muestreo puede estar constituido por los siguientes elementos: archivo clínico, archivo del servicio, libreta de citas, listado de seguros de vida, censos, listado de los individuos de una población, de hospitales o directamente de la población.

De acuerdo con las características del marco de muestreo, hay varios métodos para llegar a la unidad última de muestreo. Los más usados son el muestreo simple, el estratificado, el polietápico y el polietápico estratificado.

El método simple se usa cuando el marco permite llegar directamente a las unidades últimas de muestreo. Si se hace una clasificación previa de las unidades, es estratificado. El polietápico es aquel en el que se llega a la unidad última de muestreo mediante realizar varios muestreos simples en forma secuencial. El polietápico estratificado es una combinación de ambos (véase capítulo 19).

### Variables y escalas de medición

Son las características medibles en los elementos de estudio que se pueden describir de acuerdo con diferentes clasificaciones. Es factible clasificar las variables desde el punto de vista metodológico, por la relación que guardan entre sí; la variable considerada “causa” es la independiente y el “efecto” es la variable dependiente.

Desde el punto de vista del nivel de medición, las variables se clasifican en cuantitativas y cualitativas. Las cualitativas son las que miden las modalidades de una caracte-



rística. Para su medición se usan las escalas nominal y ordinal; se pueden representar en diagrama de barras y diagrama de sectores.

Las cuantitativas expresan la magnitud de alguna propiedad; se dividen en discretas y continuas. La variable discreta es la que se mide de unidad en unidad. La variable continua es aquella en la cual entre un valor y otro existe un número infinito de valores; ambas se pueden representar en histogramas y polígono de frecuencias acumuladas.

Es recomendable estudiar tantas variables como sea necesario y el menor número posible. Al seleccionar cada una de ellas se debe tener clara su relación con las hipótesis y los objetivos del estudio.

Cuando las variables resultan complejas tanto en su medición como en su descripción, es conveniente definir las operacionalmente (véase capítulo 9).

### Procedimientos de recolección de información

En esta sección se debe especificar la forma en que se obtendrán las variables definidas en el proyecto. El procedimiento de recolección de información debe ser congruente con el problema a investigar y con el diseño de estudio elegido (figura 3-2). En primer lugar se debe señalar si el método de recolección será directo o indirecto.

Se considera indirecto cuando las variables se obtienen de fuentes secundarias, como expedientes clínicos, documentos, base de datos, archivos clínicos, histológicos o radiológicos y así por el estilo, es decir, información ya existente antes de iniciar la investigación.

El método es directo cuando la información se obtiene de fuentes primarias, esto es, cuando es planeada por el investigador y es obtenida de los individuos involucrados en el estudio.

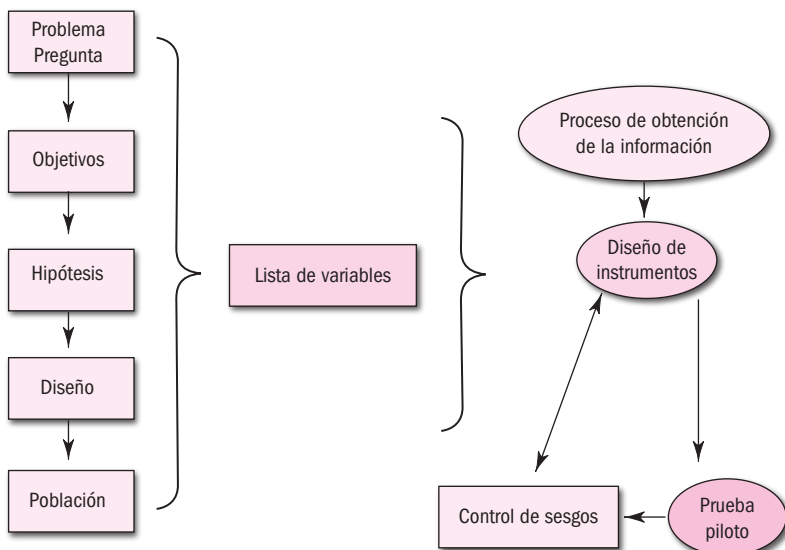


**Figura 3-2.** Proceso de obtención de la información.

En el método directo es necesario definir las técnicas de obtención, que pueden ser la observación, la exploración física, la medición, la encuesta, la entrevista, etcétera.

Las técnicas, equipo e instrumentos utilizados deben describirse con el detalle suficiente para que otros autores puedan reproducirlos. Además, es fundamental señalar los criterios de validez y control de calidad.

Cuando la obtención de información precede de fuentes secundarias, es indispensable especificar con detalle las características que deben tener.

En cada caso es importante anexar un ejemplar del instrumento que se utilizará para recolectar la información, a saber, el cuestionario, la guía de observación, cédula de entrevista, hoja de captación de datos, etcétera.

- Métodos.
- Fuentes.
- Técnicas.
- Instrumentos.
- Formas de medir.

### Reglas generales en el proceso de obtención de la información

La información debe obtenerse de la manera más sencilla posible, sin que ello implique que debe ser rigurosamente planeado el proceso para su obtención.

Además, la información debe ser “medible” a partir de los métodos más precisos y exactos disponibles.

Otro factor fundamental es que la información sea captada en forma homogénea, a fin de que sea confiable (válida, precisa, exacta y repetible).

El cuadro 3-6 presenta algunas definiciones sobre conceptos que permiten el paso del “Planteamiento del problema” a la “Selección de técnicas y diseño de instrumentos de captación de la información”.

#### **Cuadro 3-6.** Definiciones sobre conceptos que permiten el paso del “Planteamiento del problema” a la “Selección de técnicas y diseño de instrumentos de captación de la información”

- Fenómeno. Evento que sucede en la Naturaleza, en forma independiente o dependiente de nuestra voluntad.
- Observación. Parte medible y registrable de un fenómeno.
- Hecho. Observación “significativa” de un fenómeno (en este caso, **significativo** quiere decir que el hecho tiene un lugar dentro del marco explicativo del fenómeno).
- Dato. Observación de un hecho, registrado en una escala de medición, sea ésta cualitativa o cuantitativa. Un “registro” es, entonces, la expresión simbólica o numérica de un solo dato, clasificado de acuerdo con una determinada escala de medición.

El concepto de **medición**, de acuerdo con las definiciones anteriores, adquiere una gran importancia en la investigación cuantitativa. La medición es el proceso de asignación de símbolos o valores a la observación realizada de un hecho significativo para la investigación (véase capítulo 9).

#### Fases en el proceso de medición

- Comparación de una observación con una “regla estándar”, común para los investigadores.
- Asignación de un símbolo o valor uniforme y homogéneo a la observación, dentro de un sistema de clases.
- Comparación del símbolo ya clasificado contra un sistema de valores, aceptado para la mayoría de los investigadores.

### Organización y análisis de datos

El proceso de organización y análisis de la información constituye la fase en la cual los datos —u observaciones expresadas en forma de símbolos— son organizados, descritos y comparados entre sí. Por lo general la descripción y la comparación o análisis se realizan a través de técnicas estadísticas. La comparación se realiza después de la organización y la descripción, cuando los datos fueron ya clasificados según algún tipo especial de “acuerdo” o convenio.

En prácticamente todos los casos, esta clasificación obedece a un ordenamiento lógico, biológico o social de las variables que tienen importancia en la investigación. En algunas circunstancias no se requiere clasificación. Los resultados son, de acuerdo con los anteriores puntos de vista, la expresión estadística concreta que se obtiene después del análisis de los datos (véase capítulo 17).

#### Presentación de la información

Debe elaborarse el modelo de los cuadros y gráficas que se consideren necesarios para mostrar la información; asimismo, resulta de utilidad listar el título completo de cada uno de ellos, sin perder de vista que los objetivos y las hipótesis del estudio son la guía, y tomando en cuenta el tipo de variables y escalas de medición para seleccionar la manera más adecuada de presentar la información.

#### Análisis de datos

Es la forma en que serán manejadas las variables para su estudio; se debe especificar la manera en que se realizará ese procedimiento. Existen diversos métodos de procesamiento que permiten analizarlas con mayor facilidad. En un inicio se deben señalar las medidas de resumen, tanto las de dispersión como las de tendencia central. Posteriormente se definirán las pruebas estadísticas. Dentro de éstas, las más utilizadas y sencillas son: la

chi cuadrada, útil para comparar proporciones entre dos o más grupos; la *t* de Student cuando se desea comparar promedios entre más de dos grupos; el análisis de varianza, utilizado para comparar promedios entre más de dos grupos; el coeficiente de correlación, para determinar el grado de asociación entre dos variables. No obstante, dependiendo de la pregunta de investigación, existe una amplia gama de posibilidades para analizar la información. En el proyecto no es suficiente sólo señalar el tipo de medidas o pruebas estadísticas a utilizar; se deben mencionar además las variables que se describirán y analizarán (véase capítulo 17).

### Lineamientos éticos

En esta sección se explican con detalle los riesgos y peligros de la investigación cuando se utilicen radiaciones ionizantes y electromagnéticas, isótopos radiactivos, microorganismos patógenos o material biológico que los contenga, ácidos nucleicos recombinantes, así como las medidas que deban tomarse para evitar procedimientos inadecuados.

En toda investigación con seres humanos, se debe especificar la manera en que se cumplirán los preceptos éticos. El paciente y los familiares deberán ser informados de su participación en la investigación y en los casos que se considere necesario deberá ser recabada su conformidad por escrito en la **carta de consentimiento** elaborada para ese fin (véase capítulo 7). El consentimiento debe contemplar los puntos que se señalan en el cuadro 3-7.

## III. Implementación

### Recursos y logística

Es preciso estipular todos los elementos necesarios para el desarrollo del estudio; deben señalarse los recursos humanos, materiales y financieros, así como el uso que de ellos se hará a lo largo de la investigación.

**Cuadro 3-7.** Contenido de la carta de consentimiento informado

|                                                             |
|-------------------------------------------------------------|
| • Forma de participación en el estudio                      |
| • Alternativas de tratamientos conocidas                    |
| • Si se realizará asignación aleatoria a distintos grupos   |
| • Si se realizará cegamiento                                |
| • Efectos secundarios potenciales                           |
| • Derecho de no participar                                  |
| • Derecho de abandonar el estudio                           |
| • Derecho a recibir el mismo trato en caso de no participar |
| • Confidencialidad                                          |
| • Autorización para analizar la información obtenida        |

**Recursos humanos.** Personal técnico o administrativo (técnicos de laboratorio, encuestadores, enfermeras, secretarías, profesionales u otros) que participen en la investigación.

**Recursos materiales.** Descripción en cantidades de todos los elementos necesarios para la realización del estudio (reactivos, material y equipo de laboratorio y gabinete, archivos, expedientes, camas de hospital, transporte, etc.).

Se deben especificar los elementos con los que se cuenta, así como los que deberán ser adquiridos, con indicación del orden previsto para hacerlo; de manera complementaria es necesario señalar los departamentos, servicios, unidades profesionales o paraprofesionales necesarios en la investigación.

**Recursos financieros.** Costos totales de la investigación (salarios, materiales, equipo, copias fotostáticas, etc.). Si se tiene, ha de señalarse la fuente de financiamiento.

## Logística

Señala cada una de las etapas del estudio (figura 3-3).

1. Recopilación bibliográfica.
2. Elaboración del protocolo.
3. Adquisición del material.
4. Diseño de los instrumentos.

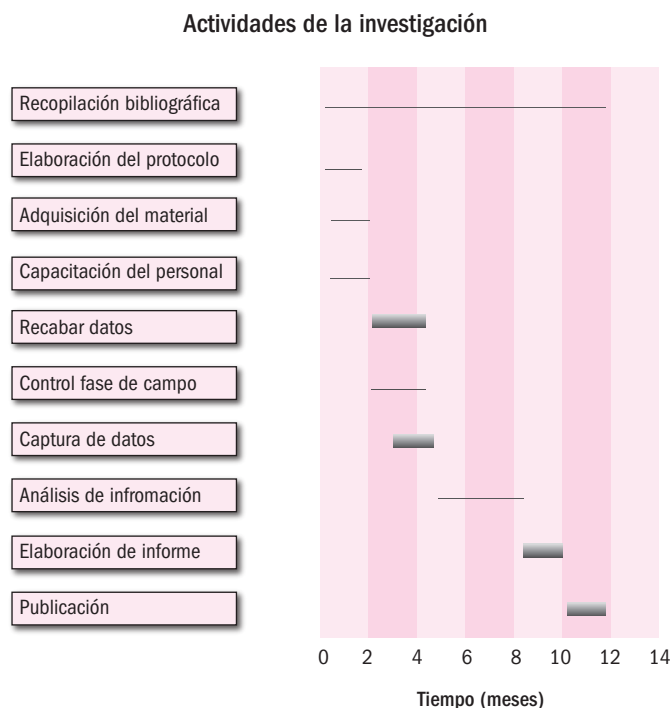


**Figura 3-3.** Gráfica de Gantt para programación logística de un proyecto de investigación.

5. Capacitación del personal.
6. Recolección de datos.
7. Procesamiento.
8. Análisis.
9. Redacción.

#### IV. Referencias bibliográficas

La información que se utiliza para construir los antecedentes y el planteamiento del problema de los proyectos de investigación se presenta como cita o referencia bibliográfica. Éstas pueden provenir de libros, revistas, índices, catálogos, documentos mimeografiados, comunicaciones personales, bases de datos en Internet, etcétera.

El propósito de las citas bibliográficas es, por un lado, otorgar los créditos a los autores correspondientes y, por otro, fundamentar el problema de investigación, así como permitir que los lectores conozcan las fuentes documentales en relación con los autores, el tipo de revista, el año en que se realizó el estudio, entre otros aspectos.

Existen distintas maneras de referir y de escribir las citas bibliográficas; la más común es la del sistema Harvard, el cual recomienda que los autores se citen con números arábigos progresivos de acuerdo con el orden de su aparición en el texto. Si existen varias referencias del mismo autor, se le asignará un número a cada una. Cuando se cita varias veces el mismo artículo se le asigna siempre el mismo número.

Finalmente, es necesario enfatizar que el éxito de una investigación depende en gran parte de la calidad y rigor metodológico con que se diseñe el protocolo.

---

#### Bibliografía

- Argimon PJM, Jiménez VJ. *Métodos de investigación clínica y epidemiológica*, 2ª ed. Harcourt, España, pp. 119-126, 2000.
- Dawson B, Trapp RG. *Bioestadística médica*, 3ª ed. El Manual Moderno, México, pp. 75-80, 2002.
- Hulley S, Cummings S, Browner W, Grady D, Hearst N, Newman T. *Designing clinical research*, 2ª ed. Lippincott Williams & Wilkins, Filadelfia, EUA, 2001.
- Méndez RI, Namihira GD, Moreno AL. *El protocolo de investigación: lineamientos para su elaboración y análisis*. Trillas, México, 1987.
- Tamayo Tamayo M. *El proceso de la investigación científica*. Limusa, México, 2001.



## Capítulo 4

### Diseños metodológicos en Epidemiología

*Laura Moreno Altamirano  
Alejandra Moreno Altamirano*

#### Criterios utilizados en los diseños metodológicos

Uno de los principales problemas de la investigación médica radica en el planteamiento de la pregunta de investigación; es decir, cuando la pregunta de investigación está mal planteada, con frecuencia las respuestas no cumplen con el nivel de cientificidad adecuado.

Varias consecuencias, generalmente interrelacionadas, resultan del planteamiento inadecuado de las preguntas de investigación. Entre las más relevantes se pueden mencionar: imprecisión en las hipótesis, ambigüedad de los objetivos del estudio, elección inapropiada del diseño metodológico, selección inadecuada de los sujetos de estudio y de las variables de interés, indefinición de los procedimientos a realizar, caracterización errónea de los resultados encontrados y extrapolación inapropiada de las conclusiones obtenidas.

Cuando la pregunta de investigación ha sido planteada de modo correcto y la fundamentación del estudio ha sido elaborada, el siguiente paso es seleccionar el diseño de investigación más conveniente (véanse caps. 2 y 3). De otra forma no será posible corregir errores durante la realización del estudio, ya que las pruebas estadísticas más complejas y sofisticadas no resolverán un diseño insuficiente.

Así, el diseño metodológico depende de manera fundamental de la pregunta de investigación que se desee contestar, de los objetivos a alcanzar y de la hipótesis a contrastar, aunque los recursos humanos, físicos, financieros y de tiempo desempeñan un papel importante.

Los diseños de investigación pueden clasificarse de acuerdo con los siguientes criterios:

1. *Interferencia del investigador:* se refiere a la que existe en la asignación de los individuos a la intervención en estudio. Esto es posible hacerlo en los estudios experimentales y cuasi-experimentales, mientras que en los observacionales no lo es.
2. *Temporalidad:* está en relación con el inicio de la ocurrencia de la enfermedad o evento de interés. En estudios retrospectivos este evento ya ocurrió, en los prospectivos no se

ha presentado y en los ambispectivos se observa una combinación de las posibilidades anteriores.

3. *Direccionalidad*: tiene que ver con la secuencia entre la exposición al factor de riesgo y el efecto. Son progresivos cuando van de la causa al efecto, regresivos de efecto a causa, o bien, de medición simultánea.
4. *Número de mediciones*: hace referencia a si se realiza una sola medición de la o las variables en estudio (diseños transversales) o si realizan dos o más mediciones (diseños longitudinales).
5. *Tipo de fuente de datos*: época y condiciones en que fue obtenida la información.

Algunos autores incluyen además el criterio de “unidad de análisis”, mismo que permite identificar los estudios ecológicos en los que se estudian conglomerados, y la unidad de análisis es un grupo (por ejemplo, un municipio o un país), mientras que en los estudios observacionales, los ensayos clínicos y los ensayos de campo, la unidad de análisis es el individuo.

Otro criterio que se utiliza con frecuencia es el número de poblaciones estudiadas; así, pueden ser estudios descriptivos o comparativos.

Hay diferencias en la nomenclatura de estos criterios, lo que ha dado lugar a confusiones importantes. En este capítulo se definirán inicialmente cinco pares de criterios, de los que se señalarán las diferencias y semejanzas mencionadas por algunos de los autores más reconocidos en el campo de la Epidemiología. Más adelante se describen los estudios clásicamente aceptados, entre los que se destacan el de casos y controles y el estudio de cohortes. El ensayo clínico controlado, por su importancia, se trata en el capítulo 5.

Es necesario enfatizar que denominar un estudio únicamente por uno de estos criterios no es conveniente, pues en un solo estudio —lo cual, de hecho, sucede con frecuencia—, llegan a combinarse varios criterios.

## Diseños experimentales, cuasi-experimentales y observacionales

De manera tradicional se han utilizado dos métodos en la investigación médica: la observación y la experimentación; tal división es plenamente aceptada por los epidemiólogos. El modelo o enfoque experimental es el más conocido y utilizado en Medicina y otras disciplinas.

Los estudios experimentales con seres humanos son una de las aportaciones trascendentes de la Epidemiología a la investigación clínica y comunitaria, sobre todo para probar la eficacia de determinadas terapéuticas farmacológicas o quirúrgicas y el efecto de programas específicos de educación para la salud.

Los estudios experimentales se caracterizan porque el investigador interfiere en el fenómeno que estudia. Lilienfeld los define como aquellos en donde es posible controlar las condiciones bajo las que se conduce el estudio; McMahon los llama “estudios de aplicación”, en donde se prueban maniobras o programas orientados a la prevención de una enfermedad o a la modificación de un hábito o actitud. Mausner afirma que en



un experimento el investigador estudia el efecto que produce la modificación de cierto factor, que está bajo su control.

Para Rothman, la experimentación es una observación controlada, en donde por lo menos una variable es manipulada por el investigador para conocer el efecto que tal manipulación produce sobre el resultado.

Last define los **estudios experimentales** como aquellos en donde el investigador controla directamente las condiciones del estudio.

Hennekens llama **estudios de intervención** a los experimentos o ensayos clínicos.

Kleinbaum, además de contemplar la manipulación de la variable de interés, incorpora la asignación aleatoria de los individuos investigados a los grupos de estudio, mismos que por lo menos son dos (el de estudio y el grupo control). Lo anterior permite controlar los sesgos de selección e incrementar la validez del estudio, requisito indispensable para conocer el efecto de la maniobra evaluada.

Los estudios experimentales que cumplen estos dos criterios; manipulación de la variable de interés y asignación aleatoria, son los que proveen la evidencia más confiable dentro de la investigación epidemiológica. A estos estudios se les conoce también como **ensayos clínicos controlados o aleatorizados**, y son los que permiten concluir con mayor certeza sobre las relaciones causa-efecto. En Medicina se pueden dividir en experimentos de laboratorio, ensayos clínicos controlados y de intervención comunitaria. Dentro de las limitaciones de estos estudios están los aspectos éticos y la complejidad de su conducción, por lo que se tratarán más detalladamente en un capítulo especial (véase capítulo 5).

Los diseños cuasi-experimentales son los estudios en los que no es posible hacer la asignación aleatoria, pero sí se realiza la manipulación de la variable de interés. También puede adoptarse esa modalidad en estudios de laboratorio o clínicos, y en aquellos que miden el impacto de programas o políticas de salud (cuadro 4-1).

Los estudios observacionales, en los que el investigador no interfiere con el fenómeno estudiado, son los más usados en Epidemiología. Con el objeto de contar con elementos cada vez más confiables sobre la relación causa-efecto, estos estudios han tenido un gran desarrollo.

Aunque el método epidemiológico tiene la ventaja de ser flexible, esta misma característica posee la desventaja de falta de consenso. Se han propuesto diferentes clasificaciones y nomenclatura para la investigación observacional, todas ellas con diferente nivel de aceptación.

**Cuadro 4-1.** Diseños epidemiológicos

| Estudio               | Control de la variable de intervención | Asignación aleatoria a los diferentes grupos de intervención |
|-----------------------|----------------------------------------|--------------------------------------------------------------|
| a) Experimental       | SÍ                                     | SÍ                                                           |
| b) Cuasi-experimental | SÍ                                     | NO                                                           |
| c) Observacional      | NO                                     | NO                                                           |

**Cuadro 4-2.** Diseños epidemiológicos según Kleinbaum**Estudios observacionales**

De acuerdo con Kleinbaum hay 15 diseños  
Básicos 3, híbridos 8 e incompletos 4

**Básicos**

Transversales, casos y controles y cohorte

**Híbridos**

Combina elementos de dos diseños básicos  
Extiende las estrategias de un diseño básico (repetición)  
Combina elementos de un diseño básico con elementos de un estudio no observacional

**Incompletos**

No cuentan con información de uno o más factores relevantes, estudios ecológicos

De acuerdo con McMahon, Hennekens y Lilienfeld, los estudios observacionales se agrupan en dos diseños básicos: 1) casos y controles, y 2) cohorte. Kleinbaum considera tres: 1) transversales, 2) casos y controles y 3) cohorte. Además incorpora los estudios híbridos, que son la combinación de dos diseños básicos, o bien, la extensión de las estrategias de uno básico, o la repetición de uno de ellos (como encuestas repetidas), o la combinación de un diseño básico con elementos de un estudio no observacional. En total son 15 los diseños propuestos por Kleinbaum. Uno de los más conocidos es el estudio de casos y controles anidado en una cohorte (cuadro 4-2).

Otros autores, como por ejemplo Méndez, proponen ocho diseños diferentes. McMahon menciona que hay investigaciones que desafían dicha clasificación, lo que ha generado múltiples interpretaciones y aun equívocos.

## Diseños retrospectivos y prospectivos

Este criterio de temporalidad tiene más de dos interpretaciones. Un grupo de autores los clasifica de acuerdo con la ocurrencia de la enfermedad o del evento resultado de interés. Si éste ocurre antes del inicio del estudio y se procede a reconstruir el evento en el pasado, se denominan **retrospectivos**, como son el estudio de casos y controles con casos prevalentes y los estudios de cohorte histórico (perspectiva histórica). Por el contrario, si el evento o la enfermedad ocurre después del inicio del estudio, se les denomina **prospectivos**; tal es el caso de los estudios de cohorte, los ensayos clínicos y los comunitarios. Ahora bien, considerando este criterio, un estudio de casos y controles con casos incidentes es un estudio prospectivo debido a que el “evento resultado” no ha ocurrido al inicio del estudio.

Otros autores los definen de acuerdo con la época de captación de la información. Si ésta se obtuvo en el pasado con fines ajenos a la investigación, se denominan **estudios retrospectivos**. Cuando la información se obtiene en el futuro y esto se planeó a propósito de la investigación, se denominan **estudios prospectivos**.

**Cuadro 4-3.** Estudios prospectivos y retrospectivos de acuerdo con diferentes autores. Época de captación de la información

| Autor                                                  | Estudios prospectivos                                                                                            | Estudios retrospectivos                                                                                                               |
|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| McMahon, Méndez, Goldstein                             | Información planeada para fines de la investigación: fuente directa de los individuos involucrados en el estudio | Información captada en el pasado con fines ajenos a la investigación: fuente, bases de datos, archivos, censos, expedientes, etcétera |
| Last                                                   | Ambos enfoques                                                                                                   | Ambos enfoques                                                                                                                        |
| Lilienfeld, Mausner, Abramson, Schlesselman, Kleinbaum | El evento de interés o enfermedad ocurre después del inicio de la investigación                                  | El evento de interés o enfermedad ocurre antes del inicio de la investigación                                                         |
| Hennekens                                              | Cuando aún no se presenta el efecto                                                                              | Cuando ya se presentó el efecto                                                                                                       |

Lilienfeld, Mausner, Abramson, Schlesselman y Kleinbaum los dividen en retrospectivos y prospectivos, considerando prospectivos a los estudios de causa-efecto y retrospectivos a los de efecto-causa; en este sentido es preferible incluir el criterio de direccionalidad.

Hennekens, por su parte, considera que los enfoques ya señalados se prestan a confusión, así que propone que los estudios se definan de acuerdo con la presencia o no del efecto buscado, en el momento de iniciar el estudio (cuadro 4-3).

## Direccionalidad

De acuerdo con este criterio, que se refiere a la forma de hacer la investigación, los estudios se clasifican en **progresivo**, **regresivo** y **de medición simultánea**.

Los estudios **progresivos** son los que van de la causa al efecto; en estos estudios la selección de la población se hace según su estatus de exposición (causa) en individuos libres de la enfermedad. Después de un periodo de seguimiento, el cual es establecido por el investigador, se registra la enfermedad o el resultado, como es el caso de los estudios de cohorte.

Si el criterio para incluir a los sujetos es la presencia o ausencia de la enfermedad, y posteriormente se recopila información sobre la exposición ocurrida en el pasado, se considera un estudio de efecto-causa y se denomina **regresivo**, como son los diseños de casos y controles.

En los estudios de **medición simultánea** tanto la exposición como la enfermedad o resultado se miden al mismo tiempo; por tanto, no hay ningún criterio (exposición o enfermedad) para selección a los sujetos que conformarán la muestra en estudio. Es decir, como ocurre en los estudios transversales, lo fundamental es que la muestra sea representativa de la población de la que provienen (figura 4-1).

| Diseños epidemiológicos |                                  |                     |
|-------------------------|----------------------------------|---------------------|
| Direccionalidad         | (Kleinbaum, Kupper, Morgenstern) |                     |
| Progresivo              | Regresivo                        | Medición simultánea |
| Causa → efecto          | Efecto → causa                   | No direccional      |
| Cohorte                 | Casos y controles                | Transversal         |

**Figura 4-1.** Diseños epidemiológicos. Direccionalidad (Kleinbaum, Kupper, Morgenstern).

Diseños transversales y longitudinales

Esta clasificación se define con base en número de observaciones o mediciones que se hacen por unidades de estudio o individuos a lo largo del tiempo.

Los estudios transversales se caracterizan porque las variables involucradas se miden en una sola ocasión. A estos estudios se les ha llamado **de corte transversal** (según Hennekens y Kleinbaum), **estudios instantáneos** (de acuerdo con Abramson), **verticales** (según McMahon) y **transversales** (según Lilienfeld).

Casi todos los autores coinciden en que el objetivo de estos estudios es conocer, en un mismo punto del tiempo, la causa y el efecto; por ejemplo, las concentraciones de colesterol y su relación con la presión arterial en una misma población y en un mismo punto en el tiempo. Se considera que el criterio contrario al transversal es el longitudinal o de seguimiento, en donde se estudia una o varias poblaciones durante un periodo. A diferencia de los estudios transversales, en los longitudinales la medición de variables se hace en dos o más ocasiones.

Los estudios longitudinales pueden ser, de acuerdo con Méndez, descriptivos o comparativos. Según este autor, son descriptivos cuando se incluye sólo una cohorte y comparativos cuando se cuenta con dos o más grupos (expuestos y no expuestos).

Es importante aclarar que si bien la encuesta o estudio transversal está conformado por un solo grupo, para establecer comparaciones y hacer inferencias sobre la prevalencia de la enfermedad, éste puede dividirse en subgrupos para su análisis, por ejemplo, sexo, grupos de edad, etcétera.

Los estudios de casos y controles son longitudinales y comparativos y, por último, los estudios de revisión de casos se consideran estudios longitudinales y descriptivos.

Diseños descriptivos o comparativos

Diversos autores, como Fox, Hall, Mausner, Bahn, McMahon, Pugh, Lilienfeld y Goldstein, entre otros, han dividido los estudios observacionales en dos categorías: descriptivos y comparativos.

Los estudios descriptivos son útiles cuando se pretende detallar la distribución y frecuencia de la enfermedad, y los comparativos cuando se desea probar una hipótesis explicativa sobre las determinantes de una enfermedad.

Algunos autores —como Méndez, Goldstein y Hennekens— consideran descriptivos a los estudios que incluyen una sola población, y comparativos a los que involucran dos o más poblaciones.

McMahon y Pugh consideran que los aspectos descriptivos de la enfermedad, como tiempo, lugar y persona, deben ser considerados un paso previo al planteamiento de una hipótesis sobre factores causales. Así que para ellos los estudios descriptivos son sólo exploratorios, indispensables cuando se desconocen las causas de una enfermedad y su distribución, pero sólo para dar paso a investigaciones posteriores de carácter comparativo o analítico.

Lilienfeld, al clasificar los estudios descriptivos, los divide en estudios de mortalidad y de morbilidad, en donde incluye aspectos de tiempo, lugar y persona.

Gran número de autores no considera “investigaciones” a los estudios de una sola población y es frecuente que les llamen “pesquisas”.

Lilienfeld afirma que, cuando se estudia la distribución de una enfermedad o fenómeno en una sola población, la asociación encontrada puede dar lugar a falacias. Para confirmar tales asociaciones, es necesario realizar estudios comparativos de manera que sea posible observar si la enfermedad o fenómeno se halla en individuos con determinada característica o sin ella.

Schlesselman les llama **estudios exploratorios** y señala que pueden ser comparativos pero son siempre transversales. Susser menciona que los estudios descriptivos son útiles para conocer, en términos numéricos, la distribución de una variable. Hennekens, por su parte, dedica un amplio espacio a los estudios descriptivos, los clasifica en estudios de correlación, informes de caso (o series de casos) y encuestas.

Kleinbaum considera que los estudios descriptivos por lo general se llevan a cabo cuando se sabe poco sobre la ocurrencia, historia natural y determinantes de la enfermedad. El objetivo de este tipo de estudios es estimar la frecuencia del fenómeno y sus tendencias en una población particular, y de ahí generar hipótesis específicas sobre etiología, tratamiento y prevención.

El mismo autor llama **estudios etiológicos** a los comparativos y afirma que son útiles cuando se sabe lo suficiente sobre una enfermedad y es posible establecer hipótesis. Los objetivos de estos estudios son identificar factores de riesgo, estimar el efecto de los factores de riesgo o del tratamiento y sugerir futuros estudios experimentales o de intervención.

Un ejemplo típico de estudio descriptivo es el de mortalidad por asma y uso de Isoproterenol en aerosol, realizado en Inglaterra y Gales entre 1959 y 1966. Después de más de un siglo de estabilidad, la mortalidad por asma aumentó tres veces en el grupo de 5 a 34 años, y siete en el de 10 a 14. Después de estudiar diversos aspectos en relación con la enfermedad, este súbito aumento se asoció fuertemente con el uso de los entonces nuevos esquemas terapéuticos a base de corticosteroides y aerosoles presurizados del tipo

del Isoproterenol. De 180 defunciones por asma, ocurridas de 1966 a 1967, 84% de los pacientes había recibido Isoproterenol y 66%, esteroides.

Como consecuencia, a partir de 1968 los aerosoles sólo podían obtenerse mediante prescripción médica. Tanto las ventas de aerosol como las muertes por asma declinaron y para 1969 la mortalidad por asma casi había llegado a su nivel anterior.

Las tendencias en mortalidad pueden ofrecer un medio sólido para plantear hipótesis. En este caso, la asociación de Isoproterenol y muerte por asma pudo haberse estudiado para comprobar esta hipótesis, analizando a todos los enfermos de asma que tomaron broncodilatadores y comparando, contra un grupo testigo, la proporción de fallecidos y no fallecidos.

Los estudios comparativos se caracterizan, según Hennekens, porque siempre existen dos o más poblaciones que se compararán con el propósito de determinar si el riesgo de enfermar es diferente entre los individuos expuestos y los no expuestos a un determinado factor.

## Diseños epidemiológicos específicos

### Estudios de casos y controles

Se considera que el enfoque moderno de los estudios de casos y controles data de 1926, cuando Lane-Clayton estudió los factores relacionados con el cáncer de mama. Sin embargo, en 1920 Broders había ya estudiado, bajo este mismo modelo, la relación entre el cáncer epidermoide y el hábito de fumar pipa.

En 1947, Schreck y Lenowitz estudiaron la relación entre cáncer de pene, ausencia de circuncisión y hábitos higiénicos deficientes.

La relación de hepatitis B con transfusión previa se demostró mediante un estudio de casos y controles, realizado por Sartwell en 1947.

En el decenio de 1950-1959 se llevaron a cabo investigaciones sobre leucemia, cáncer de vejiga, cuello uterino, mama, pulmón y estómago mediante diseños de casos y controles. Cornfield, en 1951, demostró que el riesgo relativo se puede obtener a partir de este diseño. Posteriormente, en 1954, aplicó la estandarización directa para el control de variables que causan confusión.

En 1959, Mantel y Haenszel mostraron la utilidad de estimar el riesgo relativo y aplicar la prueba de  $\chi^2$  en la búsqueda de asociaciones causales.

Otras aportaciones sobresalientes de este tipo de diseño fueron los estudios sobre los efectos adversos de medicamentos, la asociación entre tabaquismo materno y malformaciones congénitas (Kelsey, 1978); la radiación y leucemia (Linos, 1980); uso de anticonceptivos orales y desarrollo de adenoma hepatocelular (Rooks, 1979) y, por supuesto, los ya mencionados estudios sobre tabaquismo y cáncer pulmonar, ejemplos clásicos de la Epidemiología (Wynder, Graham, Levin, Goldstein, Doll y Hill) (véase cuadro 1-1).

El estudio de casos y controles es un estudio observacional. Este estudio se realiza para comparar una población que tiene el fenómeno, la enfermedad o efecto, con otra

que no lo tiene. A la primera población se le denomina **casos** y a la segunda **controles** (véase figura 11-3).

Un número cada vez más importante de investigadores prefiere denominar **testigos** o **no casos** al segundo grupo, de manera que no llegue a pensarse que este grupo está sujeto a algún control especial o de algún otro tipo. Para los autores de este capítulo, ambos términos son sinónimos.

El estudio de casos y controles cuantifica las tasas de expuestos entre los casos y los controles, y las compara. A partir del efecto (la enfermedad o fenómeno), busca las posibles causas o factores asociados, y estima el riesgo relativo a partir de la razón de productos cruzados (véase capítulo 11).

Este tipo de diseño es útil cuando se estudian enfermedades en donde el efecto se presenta mucho tiempo después de haber ocurrido la exposición, o bien para estudiar enfermedades poco frecuentes o de causas múltiples.

Los estudios de casos y controles ofrecen una aproximación a la relación causa-efecto, pero deben considerarse preliminares en la búsqueda de esta asociación, ya que no permiten establecer la temporalidad entre la causa y el efecto. Esta limitante no los invalida, pero obliga al investigador a definir cuidadosamente las estrategias que eviten sesgos en la obtención de datos y en la forma de elegir a las unidades de estudio.

## Definición de casos

La selección de los casos incluye dos aspectos fundamentales: la definición de los criterios diagnósticos y el modo de selección de los individuos elegibles.

La definición de lo que el investigador considerará un caso (la presencia de enfermedad en un individuo) debe ser homogénea. Por esta razón, todos los sujetos estudiados deben cumplir con los criterios clínicos, de laboratorio, de gabinete, de evolución y gravedad que el investigador considere mínimos indispensables para aseverar el diagnóstico de la enfermedad investigada.

Al declararse de manera explícita y estricta los criterios diagnósticos para establecer la entidad nosológica, el clínico permite que otros investigadores contrasten sus hallazgos con los suyos, rectificándolos o ratificándolos.

La definición homogénea de lo que el investigador considera en su investigación como un caso, también hace posible la aplicación de los resultados del estudio a la práctica profesional. Esto se debe a que la seguridad del médico clínico de que sus pacientes responderán de manera similar al grupo de casos depende de que el investigador, al clasificarlos, haya considerado por lo menos los mismos criterios que el clínico utiliza cuando diagnostica a sus pacientes.

De esta manera se logra una saludable uniformidad terminológica y una relación estrecha entre investigación y aplicación clínica.

Es frecuente que durante una investigación clínica se seleccionen los casos de entre los pacientes que visitan o se encuentran en hospitales, pues resulta más fácil y económico. No obstante, esto puede producir un sesgo en la selección, ya que quienes acuden a solicitar atención médica son diferentes a quienes no lo hacen, aun estando enfermos.

Una de las limitantes más difíciles de superar en el diseño de casos y controles es esclarecer la secuencia temporal de la relación causa-efecto. La certeza de que la exposición antecede al efecto se logra sólo con estudios prospectivos, del tipo de estudio de casos y controles con casos incidentes, cohortes o ensayos clínicos controlados. Sin embargo, con el conocimiento de la historia natural de la enfermedad es posible hacer una selección adecuada de los casos y una interpretación correcta de la temporalidad de los fenómenos investigados.

Por último, es necesario señalar que la comparabilidad de los casos con los controles es un aspecto fundamental, que debe ser tomado en cuenta en la selección. En este sentido, además de las características clínicas de los individuos en estudio, deben considerarse factores como edad, sexo, raza, nivel socioeconómico y, en general, todos aquellos otros de los que se sospeche una posible influencia sobre el fenómeno a investigar. Sólo de esta manera es posible establecer una comparación válida.

### Definición de controles (o testigos)

La definición de los controles presenta mayor dificultad que la de los casos. En general, puede decirse que un control es aquel individuo que posee características similares a las que se han considerado para los casos, pero con la diferencia de que el control no presenta el fenómeno a investigar (la enfermedad).

Aunque no hay un tipo de grupo control óptimo para todas las situaciones, debe siempre existir la evidencia de que los controles no tienen la enfermedad en cuestión; además, han de presentar la misma probabilidad que los casos en cuanto a estar expuestos al factor de riesgo estudiado.

Así, por ejemplo, al estudiar la relación entre infarto al miocardio e ingestión de ácido acetilsalicílico (aspirina), no serían controles adecuados los pacientes con artritis reumatoide o úlcera péptica, ya que la exposición al ácido acetilsalicílico es mayor en los pacientes con artritis reumatoide y menor en aquellos con úlcera péptica. En otro ejemplo, al estudiar sobre cáncer de cuello uterino y uso de anticonceptivos orales, en el grupo control no podría incluirse a mujeres que han usado métodos de barrera para el control de la natalidad, métodos que se considera disminuyen el riesgo de cáncer.

En general, los controles se deben seleccionar del mismo hospital o unidad de donde provienen los casos. Aunque seleccionarlos de los hospitales resulta más rápido y económico, si se obtienen de la comunidad se evitan los posibles sesgos mencionados.

Los controles, por otra parte, deben ser representativos de la población de la que se seleccionaron los casos, y no de la población que no presenta enfermedad. De esta manera, la única diferencia inicial entre casos y controles será la presencia o ausencia de enfermedad. La diferencia real, que en pocas palabras es lo que el estudio busca identificar, será la tasa de exposición al riesgo que se presume asociado con el fenómeno patológico.

Para lograr todo lo anterior, es necesario aplicar a los controles idénticos criterios de inclusión y exclusión que los ya aplicados a los casos. Cuando se seleccionan los controles de pacientes hospitalizados, debe tomarse en cuenta que un mismo factor de riesgo puede estar



asociado con diversas enfermedades, seleccionando individuos con padecimientos que no involucren directamente los factores de estudio.

Al elegir los controles entre miembros de la población abierta, se debe considerar que los individuos sanos no aceptan participar fácilmente en los estudios, y cuando lo hacen no recuerdan tan fácilmente las características de la exposición, ya que no han sufrido una enfermedad que los obligue a recordarla.

Otra fuente utilizada son amigos, esposos, parientes o vecinos de los mismos casos. Empero, es frecuente que el factor de riesgo estudiado sea, en estos grupos, similar al de los casos, subestimándose el verdadero efecto de la exposición.

Debido a que en todas las fuentes hay ventajas y desventajas, algunos autores proponen seleccionar más de un control por cada caso estudiado.

## Método de selección

En la selección de las unidades de estudio lo más conveniente es realizar un muestreo aleatorio, tanto de casos como de controles, de manera que todos los individuos elegibles tengan la misma probabilidad de ser incluidos en la investigación. No obstante, debido a que el estudio de casos y controles en muchas ocasiones se lleva a cabo en enfermedades poco frecuentes, conviene incluir a todos los pacientes elegibles. El muestreo aleatorio se utiliza sobre todo para seleccionar a los controles. Esta técnica de muestreo tiene como finalidad disminuir o evitar los sesgos en la selección de sujetos.

En especial se debe cuidar la sobrerrepresentación o subrepresentación, tanto de casos como de controles.

Las técnicas de muestreo más utilizadas comprenden muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y apareamiento.

En el capítulo 11 se encuentra mayor información sobre asociación, sesgo, exposición y técnicas de análisis de datos. La estimación de la exposición se puede tratar básicamente de tres maneras: como presente o ausente, en grados y continua.

El cuadro 4-4 muestra algunas de las ventajas y desventajas de los estudios de casos y controles.

## Estudios de cohorte

Dentro de los estudios observacionales, los estudios de cohorte ofrecen los datos más confiables sobre la relación causa-efecto. Surgen como alternativa a los estudios experimentales, en donde en muchas ocasiones los aspectos éticos no permiten la experimentación.

Desde principios del siglo xx se ha desarrollado este tipo de estudios. El de Goldberger, de 1914, en el cual investigó la relación entre factores económicos e incidencia de pelagra, fue uno de los primeros estudios de cohorte. En 1955, Doll publicó el trabajo *Mortalidad por cáncer de pulmón en los trabajadores del asbesto*. El Estudio Framingham, realizado por Dawber y colaboradores en 1959, se publicó después del seguimiento de cohortes durante seis años. En 1951, Doll y Hill iniciaron el estudio de mortalidad en

**Cuadro 4-4.** Ventajas y limitaciones de los estudios de casos y controles

| Ventajas                                                                              | Limitaciones                                                                                                      |
|---------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Útil en enfermedades raras o de periodo de latencia largo                             | No permiten estudiar enfermedades asociadas con exposición poco frecuente                                         |
| Pueden explorarse simultáneamente varios factores de riesgo para una misma enfermedad | No es posible calcular directamente la tasa de incidencia entre los expuestos y los no expuestos                  |
| Requieren menor tamaño de la muestra que otros diseños                                | Es difícil establecer la relación causa-efecto                                                                    |
| No exponen a riesgo alguno a sujetos estudiados                                       | Constituyen el diseño en el que se puede presentar mayor número de sesgos, tanto de selección como de información |
| Son relativamente rápidos y baratos                                                   | La selección de los controles es difícil                                                                          |
|                                                                                       | La información es defectuosa, ya que se obtiene de archivos y casi nunca se puede validar                         |

relación con el hábito de fumar, cuyo seguimiento duró 10 años. Sheridan, en 1950, inició un estudio en niños cuyas madres tuvieron rubéola al principio del embarazo. En 1966, Bizzozero publicó el estudio de leucemia relacionada con la radiación en Hiroshima y Nagasaki producida por la bomba atómica en 1946. Entre 1956 y 1960, Morris estudió la incidencia de cardiopatía isquémica en empleados de autobuses de Londres. Actualmente, los estudios de cohorte siguen siendo realizados con éxito por los investigadores más prestigiados (véase cuadro 1-1).

El estudio de cohorte es un estudio observacional; se considera que el estudio es de cohorte retrospectivo cuando el efecto ya se presentó, y de cohorte prospectivo cuando aún no ha ocurrido.

En este apartado se hace referencia a los estudios de cohorte prospectivos, en donde un grupo está expuesto al factor en estudio y otro no, buscando el efecto o enfermedad (véase figura 11-2).

En este diseño debe existir toda la evidencia de que la enfermedad en cuestión no existe al inicio del estudio y de que la mayoría de los individuos podrá estudiarse durante el tiempo necesario para evaluar la incidencia de la enfermedad.

Con este diseño, sí es posible medir la relación temporal entre la exposición y la enfermedad, y es útil para evaluar fenómenos patológicos que presumiblemente resultan de exposiciones “poco comunes”. También es factible evaluar varios efectos relacionados con un mismo factor de riesgo.

El seguimiento prolongado por meses, años o decenios conlleva una gran pérdida de individuos durante el mismo, lo que representa el problema más importante de este diseño. Para contrarrestarlo se recomienda iniciar con muestras grandes, por lo que resultan estudios costosos y su conducción lleva mucho tiempo.

Tampoco es un estudio útil para enfermedades poco frecuentes debido a que su desarrollo puede tardar mucho más tiempo del calculado, no presentarse con la magnitud que permita un análisis estadístico o simplemente no aparecer.

**Cuadro 4-5.** Ventajas y limitaciones de los estudios de cohorte

| Ventajas                                                                  | Limitaciones                                                                |
|---------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| Son útiles para exposiciones diversas                                     | Son poco útiles para enfermedades poco frecuentes                           |
| Pueden evaluar múltiples efectos de una misma exposición                  | Pueden ser muy costosos y tardados                                          |
| Se presentan pocos sesgos por ser estudios prospectivos                   | La validez de los resultados quizá se afecte por pérdidas en el seguimiento |
| Permiten medir la incidencia de la enfermedad en expuestos y no expuestos |                                                                             |

Se recomienda llevar a cabo un estudio de cohortes después de probar una hipótesis, por medio de un diseño de casos y controles. Ambos estudios tienen sus ventajas y sus limitaciones, por lo que se debe seleccionar el diseño más adecuado en relación con objetivos del estudio, hipótesis y cantidad de conocimiento existente sobre el problema (cuadro 4-5).

### Selección del grupo expuesto

La selección de este grupo obedece al tipo de exposición estudiada. Si se trata de alguna muy frecuente, como alcoholismo o tabaquismo, la selección es relativamente fácil. En cambio, cuando se trata de un factor poco frecuente, como algún factor ambiental limitado a zonas geográficas específicas, entonces la selección es más complicada. Ello tiene, sin embargo, la ventaja de que los niveles de exposición son más homogéneos en la población elegida como cohorte expuesta.

Incluir individuos con una exposición especial permite no sólo estudiar enfermedades de exposición infrecuente en la población general, sino disminuir el tamaño de la muestra. Por ejemplo, el angiosarcoma hepático, enfermedad poco común, se presenta aproximadamente en uno de cada 500 000 habitantes. Si en un estudio de cohorte se incluyeran 10 000 individuos de la población general, lo más probable es que no se encuentre ningún caso en varios años. Si el estudio se llevara a cabo en trabajadores expuestos a cloruro de vinilo, entre quienes la tasa de enfermedad es varias veces más alta que en la población general, la muestra podría ser mucho menor que la primera.

Cuando se desea evaluar varios efectos en un grupo de sujetos, lo ideal es que se seleccionen de una población abierta. Como es evidente, la selección del grupo expuesto depende de la hipótesis y de los objetivos de estudio.

Hay estudios en donde la exposición que se desea evaluar es de diferente grado, lo que obliga a dividir alguna cohorte en subgrupos, ya sea desde la planeación del estudio o bien durante el análisis de los datos. En ambas circunstancias es necesario establecer claramente los criterios de subdivisión de la cohorte.

A fin de obtener la información adecuada en relación con la exposición, por lo general se requiere del uso de varias fuentes. Los métodos o técnicas para definir esta exposición dependen del factor de riesgo estudiado. En ocasiones es posible documentar la exposición

a través de expedientes clínicos o laborales con la ventaja de ahorrar tiempo y dinero, pero con la desventaja de desconocer la calidad de la información.

La información documental por lo general es insuficiente o inadecuada; esto obliga a realizar entrevistas o aplicar cuestionarios a los individuos incluidos en el estudio. Esta información es mucho más confiable, ya que se recolecta por medio de técnicas y criterios homogéneos.

Al aplicarse estos instrumentos, sin embargo, pueden producirse sesgos de memoria u omisiones voluntarias de información, sobre todo en enfermedades con estigmas sociales, como el alcoholismo y el tabaquismo, o cuando implican información sobre la vida privada, como sucede con las prácticas sexuales. En otras ocasiones, la información es alterada por el sujeto de estudio al proporcionar respuestas que él considera que agradan al encuestador.

En ocasiones es necesario medir la exposición a partir de la cuantificación de elementos en el ambiente, lo cual no siempre es fácil o factible, ya que se requiere de personal, equipo y técnicas especializadas. También siguiendo este método es posible la ocurrencia de sesgos, ya que la exposición pudo producirse antes de iniciar el estudio, o modificarse cuando se efectúan las mediciones. En estos casos, una alternativa consiste en complementar la medición directa por medio de documentos.

En algunos estudios de cohorte, cuando la exposición se produce al mismo tiempo que el seguimiento, es obligatorio identificar las modificaciones de importancia que se puedan presentar como cambios de trabajo o domicilio, modificación de hábitos, tratamientos médicos, embarazo, etc. Por ello resulta necesario reevaluar de manera periódica la cohorte y valorar, de acuerdo con los criterios de selección, la permanencia o no de cada sujeto de estudio en la investigación.

La elección de las fuentes de exposición tiene su indicación, con sus propias ventajas y desventajas. El investigador debe analizar cada una de ellas y planear las estrategias que permitan tener la información más válida posible que, a pesar de las dificultades que estos estudios conllevan, la experiencia demuestra que es factible obtener.

## Selección del grupo no expuesto

La selección del grupo de no expuestos o grupo control (testigo) en el diseño de cohorte es tan importante y complicada como en el estudio de casos y controles.

Como en el caso anterior, los grupos de expuestos y no expuestos deben ser comparables; es decir, deben ser, excepto en la exposición al factor estudiado, lo más parecidos que sea posible.

Cuando la exposición se refiere a una ocupación o a condiciones ambientales, resulta difícil asegurar que el grupo control está libre de ella. En estos casos es necesario seleccionar a los no expuestos de ocupaciones o poblaciones completamente diferentes.

También ha de existir seguridad de que el grupo no expuesto pueda ser estudiado durante el tiempo suficiente para concluir en qué grado se presentó el efecto buscado.

Finalmente, es recomendable incluir varios grupos control cuando no exista la seguridad de que con un solo grupo se podrán establecer comparaciones válidas.

## Seguimiento

El seguimiento del grupo expuesto y el grupo control representa una de las mayores dificultades, ya que hacerlo de manera inadecuada puede ocasionar que los resultados no sean extrapolables a la población general.

Para evitar lo anterior es recomendable establecer de manera muy precisa la periodicidad y número de mediciones. Esto por lo general depende del tipo de enfermedad y de su periodo de latencia. Así, en las entidades agudas, el efecto de la exposición será de horas, días o semanas; para los padecimientos congénitos o los que tienen relación con el embarazo, de meses, y para las patologías crónicas como las enfermedades cardiovasculares, será de años o decenios. Nuevamente, el problema de investigación, las hipótesis aventuradas y los objetivos del estudio determinarán el proceso metodológico de acercamiento al objeto investigado.

También es necesario establecer la estrategia para mantener el contacto con las cohortes en estudio, la forma de evitar las pérdidas y la consignación de las causas de estas pérdidas. Esto último es muy importante, ya que no es lo mismo la pérdida por abandono voluntario del estudio, que por fallecimiento.

La figura 4-2 muestra una comparación entre distintos modelos.

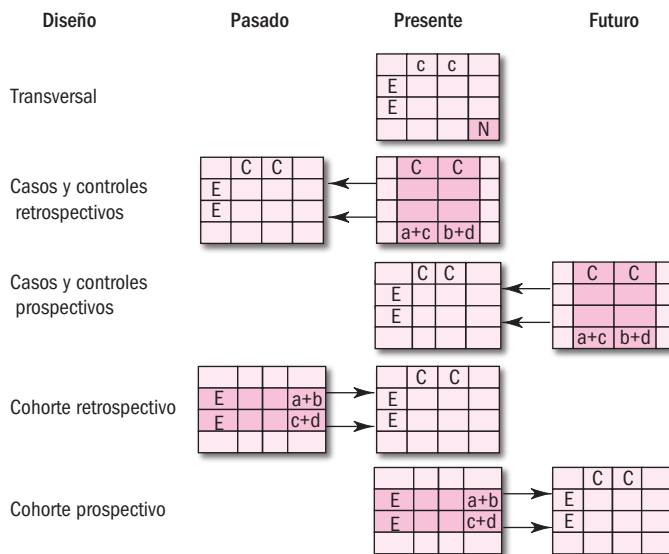


**Figura 4-2.** Diseños epidemiológicos. Comparación de diferentes modelos. C, caso; E, exposición.

## Bibliografía

- Abramson HJ. *Survey methods in community medicine*. Churchill Livingstone, Nueva York, EUA, 1979.
- Bizzozero OJ, Johnson KG. Radiation-related leukemia in Hiroshima and Nagasaki 1946-1964. Distribution, incident and appearance time. *N Engl J Med* 274 (20):1095-1101. 1966.
- Broders AC. Squamous-cell epithelioma of the lip. *J Am Med Assoc* 74:656-664. 1920.
- Cornfield J. A method of estimating comparative rates from clinical data. *J Nat Cancer Inst* 11:1269-1275. 1951.
- Dawber TR, Kannel WB, Gordon TO. Coffee and cardiovascular disease: observations from the Framingham Study. *N Engl J Med* 291:871-874. 1974.
- Doll R, Hill AB. Smoking and carcinoma of the lung. *Br Med J* 2:739-748. 1950.
- Doll R, Hill AB. A study of the etiology of carcinoma of the lung. *Br Med J* 2:1271-1286. 1952.
- Fox JP, Hall CE, Elveback LR. *Epidemiology. Man and disease*. MacMillan, Nueva York, EUA, 1970.
- Goldberger J. En: Buck C, Llopis A et al. *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. OPS, Washington, EUA, pp. 630-658. 1980.
- Goldstein H. *The design and analysis of longitudinal studies, their role in the measurement of change*. Academic Press, Nueva York, EUA, 1979.
- Hennekens HCH, Buring EJ. *Epidemiology in medicine*. Little Brown, Boston, EUA, 1987.
- Kelsey JK, Dwyer TR, Bracken MB. Maternal smoking and congenital malformations on epidemiological study. *J Epidemiol Commun Health* 32:102-107. 1978.
- Hulley S, Cummings S, Browner W, Grady D, Hearst N, Newman T. *Designing clinical research*, 2a. ed. Lippincott Williams & Wilkins, Filadelfia, EUA, 2001.
- Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiology, principles and quantitative methods*. Life Time Learning Publications Belmont, EUA, 1982.
- Lane-Clayton JE. A further report on cancer of the breast. En: Schlesselman JJ. *Case control studies: design, conduct, analysis*. Oxford University Press, Nueva York, EUA, 1982.
- Last MJA. *Dictionary of Epidemiology*. Oxford University Press, Nueva York, EUA, 1982.
- Levin ML, Goldstein IL. Cancer and tobacco smoking. Preliminary report. *J Am Med Assoc* 143:336-338. 1950.
- Lilienfeld AM, Lilienfeld D. *Foundations of Epidemiology*, 2a. ed.: Oxford University Press, Nueva York, EUA, 1979.
- Linos A, Gay JE, Orves AL. Cow-dosis radiation and leukemia. *N Engl J Med* 302:1101-1105. 1980.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Inst* 22:719-748. 1959.
- Mausner JS, Bahn AK. *Epidemiology*. Saunders, Filadelfia, EUA, 1974.
- McMahon B, Pugh TF. *Epidemiology: principles and methods*. Little Brown, Boston, EUA, 1970.
- Méndez I, Namihira D et al. *El protocolo de investigación. Lineamientos para su elaboración y análisis*, 2a. ed. Trillas, México, 1990.
- Morris J. En: Buck C, Llopis A et al. *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. OPS, Washington, EUA, pp. 744-757. 1980.

- Rooks JB, Orytl W. The cooperative liver tumor study group. Epidemiology of hepatocellular adenoma: the role of oral contraceptive use. *J Am Med Assoc* 242:644-648. 1979.
- Rothman JK. *Modern epidemiology*. Little Brown, Boston, EUA, 1986.
- Sartwell PE. Infectious hepatitis in relation to transfusion. En: Schlesselman JJ. *Case-control studies: design, conduct, analysis*. Oxford University Press, Nueva York, EUA, 1982.
- Schlesselman JJ. *Case-control studies: design, conduct, analysis*. Oxford University Press, Nueva York, EUA, 1982.
- Sheridan MD: En: Buck C, Llopis A *et al.* *Desafíos de la epidemiología. Problemas y lecturas seleccionadas*. OPS, Washington, EUA, pp. 723-731. 1980.
- Schreck R, Lenowitz H. Etiology factors in carcinoma of the penis. *Cancer Research* 7:180-187. 1947.
- Schwartz D, Flamant R, Lléllouch J. *Clinical trials*. Academic Press, Nueva York, EUA, 1980.
- Susser M. *Causal thinking in the health sciences: concepts and strategies of epidemiology*. Oxford University Press, Nueva York, EUA, 1973.
- Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *J Am Med Assoc* 143:329-338. 1950.

# Capítulo 5

## Estudios experimentales. Ensayo clínico aleatorio

*Diana Vilar-Compte  
Elizabeth Salazar Rojas*



### Introducción

Un ensayo clínico planeado y conducido de manera apropiada es un diseño de investigación poderoso para evaluar la eficacia de una intervención. Proporciona uno de los niveles más altos de evidencia.

### ¿Qué son los ensayos clínicos?

Un *experimento* es una serie de observaciones efectuadas en forma controlada por el investigador. Un *ensayo clínico* es un experimento en el que se prueba un nuevo tratamiento, procedimiento, dispositivo, etc., en humanos y utiliza técnicas de intervención. El investigador controla los factores que contribuyen a la variabilidad, sesgos, tratamiento, evaluación del desenlace y el análisis. Los ensayos clínicos son estudios prospectivos, en donde los participantes se siguen desde un momento bien definido en el tiempo (tiempo cero o basal) hasta la ocurrencia del desenlace de interés.<sup>1,2</sup>

En los ensayos clínicos debe existir un grupo en el que se pruebe la maniobra *experimental* (grupo experimental) y un grupo *control* para contrastar la eficacia de la intervención. Al inicio del estudio, los participantes del grupo *control* tienen que ser lo más parecido posible a los del grupo *experimental* en los aspectos relevantes, para que las diferencias observadas en el desenlace puedan ser atribuidas a la acción de la intervención (figura 5-1). En general, la mayor parte de las intervenciones son probadas contra el mejor tratamiento estándar para la patología en cuestión. Cuando no existe un estándar de manejo puede compararse contra un grupo de individuos sin “intervención activa”, es decir, contra un placebo o ninguna intervención.<sup>2</sup>

A diferencia de los estudios experimentales en animales, en los ensayos clínicos el investigador no puede dictaminar qué es lo que hace un sujeto, y aunque se recomien-



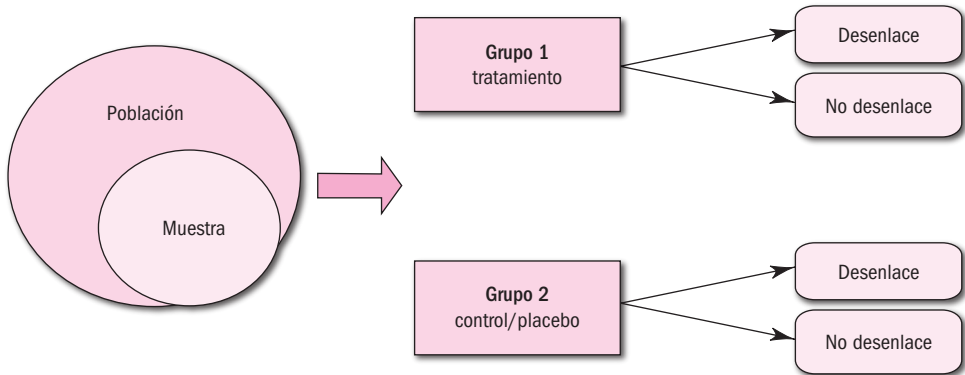


**Figura 5-1.** Características básicas de un ensayo clínico aleatorizado.

da evitar las cointervenciones (visitas a médicos fuera de la institución, automedicación, etc.), es deseable que en la planeación de un ensayo clínico se consideren las estrategias para que los participantes se apeguen lo más posible a la intervención asignada.

El ensayo clínico ideal es aleatorizado (la intervención es sorteada entre los participantes al azar, de forma tal que tengan las mismas posibilidades de recibir el tratamiento experimental o control) y doble ciego (ni los investigadores ni los participantes conocen el tratamiento que reciben [experimental o control]). De acuerdo con la Agencia de la Investigación y Calidad de Atención Médica (AHRQ, por sus siglas en inglés), cuando se realizan de una manera adecuada, proporcionan el mayor grado de evidencia para demostrar si una intervención tiene el efecto postulado (cuadro 5-1).<sup>3</sup>

En general son estudios costosos y logísticamente complejos, además de que por la naturaleza del diseño pueden tener implicaciones éticas importantes. Debido a lo anterior están reservados a preguntas de investigación relativamente “maduras”.

**Cuadro 5-1.** Clasificación del nivel de evidencia AHRQ  
(Agency for Healthcare Research and Quality)<sup>9</sup>

| Nivel | Descripción                                                                                                                                                  |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ia    | La evidencia científica procede a partir de metaanálisis de ensayos clínicos controlados aleatorizados                                                       |
| Ib    | La evidencia científica procede de al menos un ensayo clínico controlado y aleatorizado                                                                      |
| IIa   | La evidencia científica procede de al menos un estudio prospectivo controlado, bien diseñado y sin aleatorizar                                               |
| IIb   | La evidencia científica procede de al menos un estudio casi experimental, bien diseñado                                                                      |
| III   | La evidencia científica procede de estudios descriptivos no experimentales, bien diseñados como estudios comparativos, de correlación o de casos y controles |
| IV    | La evidencia científica procede de documentos u opiniones de expertos y/o experiencias clínicas de autoridades de prestigio                                  |

## Planeación del ensayo clínico

### Pregunta de investigación

La planeación de un ensayo clínico aleatorizado (ECA) depende de la pregunta que el investigador se ha planteado contestar, y se realiza de acuerdo con la variable que desea evaluar. Una pregunta bien planteada y sustentada en el conocimiento existente facilita el diseño y la conducción, por lo que ello constituye el primer paso. Como en otros diseños, puede haber más de una pregunta por responder, pero el diseño del ensayo clínico debe partir de una *pregunta principal* o *primaria* (véase cap. 3).

**Pregunta principal.** Es la que el investigador está más interesado en contestar; siempre debe ser factible y pertinente. A través de esta pregunta se efectúa el cálculo del tamaño de la muestra y se contrastan las hipótesis, considerando que el desenlace observado suele ser, en promedio, diferente en el grupo experimental con respecto al del grupo control. La variable desenlace (que en este caso es la variable dependiente principal) debe tener algún beneficio clínico relevante, como salvar la vida, disminuir los síntomas, mejorar la calidad de vida, o bien, la modificación de alguna variable “intermedia” o subrogada como la presión arterial. En ocasiones, el investigador está interesado en demostrar que no existen diferencias en el desenlace, independientemente de tratar al paciente con el fármaco experimental o control. Este último planteamiento requiere tamaños de muestra mayores y estadísticamente tiene algunos aspectos que considerar.

**Preguntas secundarias.** Se consideran “auxiliares” de la pregunta principal. El estudio debe estar diseñado para que estas preguntas puedan contestarse también, y, por tanto, tienen que contrastarse con sus respectivas hipótesis, mismas que deben especificarse desde un principio antes de empezar el estudio. Para contestar estas preguntas, las variables deben ser elegidas para que arrojen las respuestas deseadas sin perder la eficiencia en el diseño.

Existen dos grupos de preguntas secundarias; en las primeras el desenlace es diferente al de la pregunta principal. Por ejemplo, la pregunta principal evalúa si la mortalidad por infarto agudo al miocardio se modifica por la intervención, y las preguntas secundarias evalúan la incidencia de muerte por causas específicas, como la mortalidad por enfermedad coronaria o la incidencia de un infarto agudo al miocardio no fatal. El segundo grupo de preguntas secundarias está relacionado con un subgrupo de hipótesis; por ejemplo, en un estudio sobre quimioterapia para el cáncer el investigador quiere conocer específicamente el desenlace en un subgrupo de enfermos, de tal forma que comparará ese subgrupo de personas en cuestión (sujetos en el grupo experimental en ese subgrupo con el mismo subgrupo de pacientes en el grupo control). Es importante señalar que esto debe especificarse previo al inicio del ensayo, basado en expectativas realistas y entendiendo que deben ser limitadas en el número.<sup>2,4</sup>

### Selección de la población de estudio

La definición de la población de estudio es una parte esencial del planteamiento de la pregunta principal y constituye un paso crítico en el diseño del estudio. Lo anterior se

logra a través de los criterios de selección de los participantes, y éstos son los que permitirán contestar la pregunta. Al aplicar los resultados del ensayo en cuestión, el clínico debe saber con precisión cómo identificar a los pacientes que pueden beneficiarse de un mejor tratamiento.<sup>5</sup>

Para elegir a los sujetos que participarán en el estudio, se utilizan los **criterios de inclusión y exclusión**, cuyo propósito estriba en identificar a la población en la que es factible, ético y relevante estudiar el impacto de la intervención.

Los **criterios de inclusión** se definen como las características que deben estar presentes en las unidades de estudio. Deben ser tales que permitan el reclutamiento de un número suficiente de participantes con una probabilidad elevada de desarrollar el desenlace; ello, con el propósito de alcanzar un poder adecuado y observar diferencias en el desenlace. También deben estar contruidos para que los resultados sean extrapolables a otras poblaciones; por ejemplo, si el desenlace es poco frecuente, como el cáncer de estómago, será necesario incluir pacientes con factores de riesgo para esta neoplasia, con el propósito de disminuir el tamaño de la muestra y el periodo de seguimiento; sin embargo, el limitar estos criterios disminuye la generalización de los resultados y vuelve más difícil el reclutamiento de participantes.

En el proceso de inclusión debe considerarse la estratificación (por ejemplo, de una característica en particular como edad o grupo étnico), la cual permite a los investigadores incluir pacientes con características que puedan influir sobre el efecto del tratamiento o la generalización de los resultados.<sup>4</sup>

Los **criterios de exclusión** son las características que no deben estar presentes en las unidades de estudio; deben tener una particularidad importante, ser elegidos con cuidado y buen juicio, ya que exclusiones innecesarias disminuyen la generalización de los resultados, complican el reclutamiento e incrementan el costo del proceso. Tras considerar lo anterior, existen al menos tres razones por las que la inclusión inapropiada puede oscurecer la respuesta a la pregunta de investigación:

1. Que el tratamiento experimental sea dañino (por ejemplo, riesgo incrementado de un evento adverso a la sustancia activa o riesgo inaceptable de recibir un placebo).
2. Que haya pacientes que puedan fallecer durante el periodo de estudio por otras enfermedades y que ello enmascare los resultados. Lo anterior puede ser por falta de un seguimiento completo, e invariablemente incorpora “ruido” al análisis.
3. Es poco probable que resulte eficaz el tratamiento activo, ya sea por una baja posibilidad de desarrollar el desenlace en cuestión (por ejemplo, raloxifeno para prevención de enfermedad coronaria en pacientes con bajo riesgo de esta enfermedad), o porque el paciente presenta alguna enfermedad que no responde al tratamiento o está bajo alguna terapia que interfiere con la intervención (por ejemplo, raloxifeno para la prevención de enfermedad coronaria en pacientes que utilizan estrógenos, pues estos últimos compiten con el raloxifeno).

Otras razones que deben tomarse en cuenta dentro de los criterios de exclusión son pobre apego a la intervención o al seguimiento, así como problemas prácticos para la participación en el estudio (por ejemplo, retraso mental).<sup>4,5</sup>

A la par que se eligen los criterios de selección de los participantes, es necesario considerar el tamaño de la muestra. El reclutamiento suele ser más complejo en los estudios experimentales que en los observacionales, por lo que el investigador debe asegurarse de seleccionar una muestra de una población accesible y suficientemente amplia. También es necesario contar con los recursos apropiados. Los estudios con menos participantes que lo adecuado son incapaces de demostrar efectos sustanciales, suelen ser una pérdida de tiempo e incluso poco éticos, además de producir con mucha frecuencia conclusiones erróneas.

En este proceso y previo a la aleatorización, es importante que el investigador recolecte la información necesaria en relación con las variables basales con tres objetivos principales:

1. Tener la información completa de cada participante y que, en un momento determinado, sea posible consultar otras fuentes de información importantes para la adjudicación de un evento (por ejemplo, un expediente del Seguro Social si el paciente es derechohabiente y es hospitalizado por una tromboembolia pulmonar), así como datos personales que permitan su fácil localización durante el seguimiento. Es importante señalar que estos datos son confidenciales y que, por tanto, tienen que estar adecuadamente protegidos.
2. Medición de variables conocidas como de riesgo, para la variable desenlace en cuestión o que puedan emplearse para la definición de algunos subgrupos. En este sentido es importante identificar y medir variables predictoras del desenlace (por ejemplo, tabaquismo en las parejas de los sujetos participantes en una intervención para dejar de fumar), así como el efecto que éstas pueden tener sobre la intervención en grupos pequeños como modificadores del efecto. Por ejemplo, en un estudio experimental sobre el uso de alendronato para disminuir las fracturas en mujeres con osteoporosis, se observó que este fármaco disminuía de manera significativa las fracturas en las mujeres con muy baja densidad ósea, pero no ocurría lo mismo en aquellas mujeres que no tenían este problema.<sup>6</sup>
3. Medición de los valores basales de la variable desenlace. Esto es particularmente importante cuando la variable desenlace incluye una modificación en ésta, de tal suerte que pueda observarse el cambio al final del estudio (por ejemplo, modificación de las cifras de tensión arterial por un antihipertensivo). La medición al inicio y al final tiene siempre que hacerse de la misma forma y bajo los mismos criterios para garantizar su validez. Si, por el contrario, la variable desenlace es dicotómica (presencia o no de infarto al miocardio), es prudente medir la ausencia de ésta al inicio del estudio previo a la aleatorización, ya que de estar presente, se podría no estar contestando la pregunta planteada.

Cabe señalar también que en este proceso es indispensable ser cuidadoso, ya que la maniobra de aleatorización tiene como objetivo reducir los factores basales que pueden causar confusión, por lo que no es necesario medir todas y cada una de las variables. Efectuar mediciones excesivas añade un mayor grado de complejidad al estudio e incrementa los costos.

## Selección de la intervención

**Intervención.** La selección de la intervención es crítica en el diseño de un ensayo clínico. Los investigadores deben considerar la intensidad, duración y frecuencia de la intervención que mejor balancea la eficacia y la seguridad. En este proceso es preciso tomar en cuenta la factibilidad del enmascaramiento, las intervenciones simples *vs.* múltiples (por ejemplo, diferentes dosis) y la aplicación del tratamiento en la práctica cotidiana fuera de un proyecto de investigación.

Los ensayos que prueban una sola intervención son por lo general más sencillos; sin embargo, muchas enfermedades crónicas como la infección por VIH o la insuficiencia cardíaca crónica se tratan regularmente con una combinación de tratamientos, por lo que los resultados quizá no generen conclusiones claras al respecto.

**Selección del control.** El mejor control es aquel que no recibe un tratamiento activo; sin embargo, en la actualidad son pocos los estudios en los que esto es posible. Es decir, los controles deben recibir el mejor tratamiento estándar para la patología en cuestión y de ser posible, además, recibir un *placebo* que sea indistinguible de la sustancia activa. Se ha observado que lo anterior compensa los efectos potenciales del efecto placebo de la sustancia activa (por ejemplo, a través de la autosugestión u otros mecanismos no farmacológicos) y, por tanto, las diferencias entre los grupos de tratamiento pueden atribuirse al efecto biológico de la sustancia estudiada.<sup>4</sup>

**Placebo.** Un placebo es un agente farmacológicamente inactivo que los investigadores administran a los participantes en el grupo control de un ensayo clínico. El primer ensayo clínico controlado con placebo se efectuó en el decenio de 1930 cuando se probó sanocrisina en comparación con agua destilada en pacientes con tuberculosis.<sup>7,8</sup> Desde entonces los ensayos clínicos aleatorizados con placebo han sido controvertibles, especialmente cuando se priva del tratamiento efectivo al grupo control.<sup>2,7-9</sup>

De acuerdo con Friedman, existen sólo dos situaciones en las que se justifica usar un placebo: 1) cuando no existe ningún tratamiento superior al placebo y 2) si el participante entiende con claridad que se está utilizando un placebo y que tiene las mismas probabilidades de recibir éste o la alternativa.<sup>2</sup>

Por otro lado, desde hace 50 años se ha documentado que el placebo *per se* produce mejoría (objetiva o subjetiva) en 30-40% de los pacientes en un amplia gama de entidades clínicas como el dolor, el asma, la hipertensión arterial e, incluso, el infarto agudo al miocardio.<sup>10</sup>

El debate acerca del uso apropiado del placebo en las investigaciones apareció después de haberlo utilizado en numerosos ensayos clínicos, en los cuales se empleaba placebo a pesar de la existencia de un tratamiento efectivo.<sup>7,11</sup> Estos estudios violaban los principios básicos de la Declaración de Helsinki. Además, la aparición del VIH-SIDA e innovadoras metodologías para evaluar nuevas drogas (ensayos clínicos multicéntricos, financiadores externos al país, uso de compradores-placebo, por ejemplo), sobre todo en los países de menores ingresos, provocó debates éticos por parte de los miembros de las instancias reguladoras internacionales.<sup>7</sup> Uno de los casos que suscitaron este debate y

que permitió armonizar criterios para el uso de placebo fue el estudio sobre la eficacia de la zidovudina en la disminución de la transmisión vertical del HIV, en donde a pesar de los beneficios de su empleo durante el embarazo, en un ensayo clínico en África se utilizó como control un placebo, bajo el supuesto de que el tratamiento estándar era no dar ningún tratamiento dada la escasez de recursos.<sup>12</sup>

## Variables de desenlace

La definición de los desenlaces a evaluar está íntimamente relacionada con otros componentes del diseño, con los costos y la factibilidad de responder a la pregunta de investigación. Al igual que con la pregunta de investigación, es necesario comprometerse con una variable primaria de desenlace, pero contar con otras variables incrementa la riqueza del estudio y aumenta las posibilidades de un análisis secundario.

Las mejores variables desenlace son aquellas clínicamente relevantes y en las que el tratamiento puede modificar el curso. Cuando se usan variables de tipo continuo, el tamaño de la muestra es en general menor.

En ocasiones es necesario utilizar *marcadores intermedios*, tales como la densidad ósea, los cuales están relacionados con el desenlace clínico y pueden ayudar a comprender aspectos de fisiopatología, la mejor dosificación, etc. Dada su naturaleza, la precisión en su medición es fundamental para poder evaluar el cambio. Estos marcadores también pueden considerarse como *variables subrogadas* del desenlace clínico, ya que en la medida que el tratamiento induzca un cambio en el marcador, puede modificar la variable desenlace.

## Efectos adversos

En el proceso de selección de las variables de desenlace, es indispensable incluir algunas que detecten *efectos adversos* asociados a la intervención.

Los efectos adversos asociados a un procedimiento ocurren en un rango muy amplio, y es poco probable que aquellos muy raros puedan detectarse durante la conducción de un ensayo clínico; con frecuencia se descubren una vez que la terapia en cuestión se utiliza en la población abierta.

En las fases iniciales de la investigación, cuando no se conoce en forma precisa cuáles pueden ser los efectos adversos potenciales, es preferible usar preguntas y variables de muy amplio espectro que incluyan todas las posibilidades. Es importante señalar que efectos adversos comunes, como las infecciones de las vías respiratorias superiores, no requieren preguntas específicas muy precisas; sin embargo, aquellos eventos particulares asociados a ciertos medicamentos necesitan preguntas específicas capaces de detectar el evento (por ejemplo, el uso de estatinas y rabdomiólisis requiere de preguntas específicas que incluyan aspectos tales como la miositis). Si además el proyecto está involucrado en el proceso de aprobación de un medicamento, deben seguirse aspectos regulatorios específicos contenidos en las “Guías de Buenas Prácticas Clínicas”.

## Proceso de aleatorización

La aleatorización suele producir grupos de estudio comparables entre sí con respecto a los factores de riesgo, previene el sesgo de asignación de los participantes por parte del investigador, y suele garantizar niveles de significancia estadística.

El ensayo clínico aleatorio es el estándar de oro empleado para la comparación de otros diseños, ya que por sus características es el que incurre en menos sesgos. De manera simple, el proceso de aleatorización consiste en dar al participante la misma oportunidad de asignación al grupo experimental o control, con la finalidad de comparar los tratamientos sobre las variables de desenlace de interés.<sup>2,7</sup> Se acepta que la aleatorización tiene como propósito prevenir la existencia de diferencias entre los grupos, mismas que no sean derivadas de los tratamientos que se están comparando, por lo que las diferencias observadas en la variable desenlace se deben considerar estrictamente como debidas a la maniobra bajo estudio.

El concepto de “aleatorización” fue introducido por Fisher en la agronomía; su argumento principal era que prevendría las diferencias sistemáticas de cualquier tipo (voluntarias o involuntarias por parte del investigador), por lo que este concepto es preferible sobre la asignación no probabilística (sistemática, secuencial, por facilidad o por conveniencia), a la que con tanta frecuencia se recurre. Un ejemplo de una aleatorización no probabilística es la inclusión de pacientes a un grupo ciertos días de la semana, o bien, usar los números de registro pares para asignar a los individuos a un tratamiento (por ejemplo, experimental) y los registros con números impares al grupo control.

Es importante señalar que el proceso de aleatorización tiene que ser lo suficientemente seguro para que los investigadores y otros participantes en contacto con los pacientes no puedan influir sobre la asignación, por lo que es deseable que no sean los investigadores los involucrados en aleatorizar a los participantes del estudio.

Actualmente existen diversos métodos para la aleatorización y, dependiendo de la pregunta de investigación, variable desenlace y de la complejidad del estudio, se recurrirá a uno u otro. El proceso más sencillo es el que se conoce como **aleatorización simple**, en donde los individuos se asignan equitativamente en ambos grupos. Los estudios pequeños en general no requieren maniobras más complejas de aleatorización; sin embargo, en ocasiones es necesario balancear los grupos de estudio por número de participantes (aleatorización por bloques), o por determinadas variables basales predictoras del desenlace (bloques estratificados). También existen algunas otras técnicas de aleatorización; su descripción detallada escapa a los objetivos de este texto, por lo que se invita al lector interesado a profundizar en textos *ex professo*.

## Enmascaramiento o cegamiento

Es una condición impuesta sobre un procedimiento específico para intentar guardar el conocimiento del tratamiento asignado, el curso del tratamiento u observaciones previas.<sup>13</sup>

Siempre que sea posible, el investigador debe diseñar la intervención de forma tal que ni los investigadores ni los participantes ni otras personas que tengan contacto con la

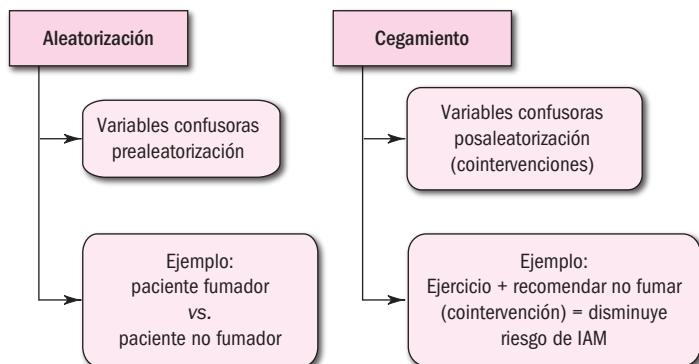


**Figura 5-2.** Estrategias de control de variables confusoras.

investigación (evaluadores del desenlace, personal de laboratorio, etc.) conozcan el grupo de asignación (experimental o control). De no ser posible enmascarar a todos, es deseable que se enmascare al mayor número de personas que estén participando en el protocolo.

El enmascaramiento es tan importante como la aleatorización; previene sesgos por otras cointervenciones, así como en la adjudicación del desenlace. Tradicionalmente se considera que la aleatorización previene la aparición de confusores presentes al momento de efectuar dicha maniobra, mientras que el enmascaramiento elimina las diferencias que aparecen entre los grupos durante el seguimiento (figura 5-2). Por ejemplo, en un ensayo no cegado, el investigador principal puede prestar más atención al grupo de pacientes que recibió el fármaco experimental, y esta simple “cointervención” puede ser la explicación de las diferencias observadas. En el caso del enmascaramiento en la adjudicación del desenlace, se observó en un ensayo clínico no aleatorizado sobre terapia hormonal de reemplazo que los investigadores eran más propensos a solicitar un ultrasonido Doppler de la extremidad en las mujeres que presentaban edema y dolor de la misma y que recibían estrógenos (fármaco experimental), comparativamente con el grupo control, en donde era menos frecuente la búsqueda intencionada de una complicación de tipo trombótico.

El enmascaramiento puede existir en tres niveles diferentes dependiendo de quiénes se encuentran cegados a la maniobra de asignación; por su naturaleza puede ser: simple, doble y triple ciego. Sus características se resumen en el cuadro 5-2.

## Diseños alternativos al ensayo clínico aleatorio y paralelo

El ensayo clínico aleatorio más conocido es aquel que involucra un grupo experimental y se compara con un grupo control; sin embargo, existen variedades dentro del diseño básico. No se pretende describir las variedades en esta revisión; sin embargo, se describen brevemente los más utilizados en la clínica. Existen textos en donde el lector puede profundizar en estas variedades si resulta de su interés.



**Cuadro 5-2.** Tipos y características del cegamiento

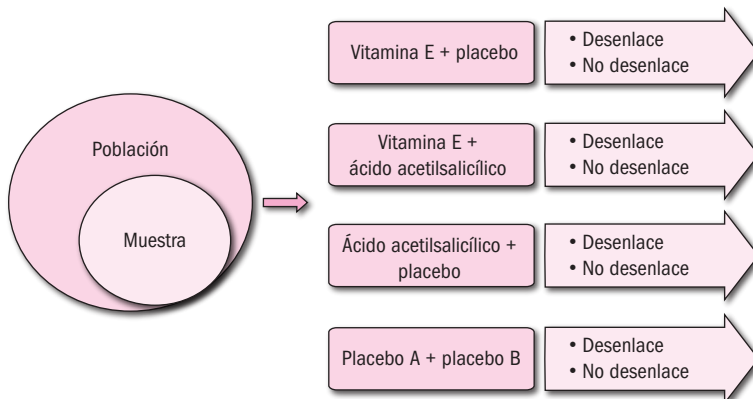
| Tipo de cegamiento | Características                                                                                          |
|--------------------|----------------------------------------------------------------------------------------------------------|
| No enmascarado     | Ensayo en el que los investigadores y participantes conocen el tratamiento asignado                      |
| Ciego simple       | Los participantes del estudio desconocen el tratamiento que reciben                                      |
| Doble ciego        | Participantes, investigadores y patrocinadores desconocen la asignación del tratamiento                  |
| Triple ciego       | Es una variante del doble ciego en la que se mantiene el enmascaramiento hasta cierto punto del análisis |

**Diseño factorial**

Este subtipo de ensayo permite contestar dos o más preguntas independientes (fármacos independientes sin ninguna interacción) a la vez en el mismo grupo de sujetos, por lo que se reconoce como un diseño eficiente. Los sujetos pueden no recibir ninguna intervención, una de ellas, o a la larga todas. Por ejemplo, en el *Women's Health Study* se probó el efecto del ácido acetilsalicílico (aspirina) en dosis bajas con la vitamina E para la prevención primaria de enfermedades cardiovasculares en mujeres sanas. Para ello se distribuyeron en cuatro grupos, como lo muestra la figura 5-3.

Se comparó la frecuencia de eventos cardiovasculares en las mujeres que sólo habían recibido ácido acetilsalicílico contra placebo y las que sólo habían recibido vitamina E contra placebo (se ignoró el hecho de que la mitad en cada grupo había recibido el otro tratamiento), de manera que los investigadores tuvieron dos respuestas en el mismo estudio.<sup>14-15</sup>

Una limitación de este diseño es que puede haber posibles interacciones entre los tratamientos, y se requieren tamaños de muestra significativamente mayores para alcanzar un poder adecuado. El reclutamiento y el apego al tratamiento pueden también ser más difíciles.<sup>7,11</sup>

**Figura 5-3.** Diseño factorial.

## Aleatorización por conglomerados o grupos

En este caso, la unidad de análisis la constituye un grupo y no un individuo. Tiene la ventaja de que en general los grupos ocurren de manera natural y se requieren muestras de menor tamaño; sin embargo, puede haber interacciones entre los individuos no evaluados y que distorsionen el desenlace evaluado. Suelen emplearse a nivel comunitario y son útiles para intervenciones educativas.

## Ensayo por equivalencia

Se realizan cuando las diferencias en tratamientos son cercanas a cero. Se ponen en práctica porque hay tratamientos que pueden diferir en seguridad, efectos adversos, costos, etcétera, y el hecho de mostrar “equivalencia” tiene importancia para el uso subsiguiente de uno o ambos tratamientos.

Tales ensayos suelen necesitar tamaños de muestra grandes para probar estas diferencias y pueden tener implicaciones éticas, particularmente si el fármaco experimental no tiene ninguna ventaja sobre el tratamiento estándar.<sup>5,7,14</sup>

## Otros. Estudios seudoaleatorizados

Este grupo de estudios no tiene una maniobra de aleatorización verdadera, por lo que la introducción de sesgos es mayor y se pierde la fortaleza del diseño experimental, la aleatorización. Un ejemplo es incluir pacientes al grupo de tratamiento de acuerdo con el día de su consulta, terminación del número de registro, preferencias del médico, etcétera. En nuestro medio esto es una práctica frecuente, y aunque tiene la ventaja potencial de facilitar la logística, tiene implicaciones en la validez de los resultados.

## Diseños cruzados

Esta variedad se usa poco; su objetivo es utilizar los mismos sujetos (individual o grupalmente) en el tratamiento experimental y control, en diferentes momentos del estudio. Por ejemplo, si el sujeto A recibe el tratamiento experimental “X” durante dos meses, posteriormente, y después de un periodo de descanso o de lavado (*washout*), se asignará al grupo control, de tal suerte que sirva como su mismo control, lo que en principio puede disminuir las diferencias al hacer la comparación (figura 5-4); sin embargo, analíticamente representan un reto y siempre puede haber un efecto residual del tratamiento previo. Se utilizan en situaciones donde la intervención sea rápidamente reversible.<sup>7,14</sup>

## Serie de tiempo

Las mediciones se realizan antes y después de que el individuo recibe la intervención, de tal forma que cada participante sirve como su propio control (figura 5-5). Se pueden prácticamente eliminar características basales innatas como grupo étnico, género, edad, etcétera; sin embargo, la falta de un grupo control concurrente es una desventaja.<sup>14</sup>

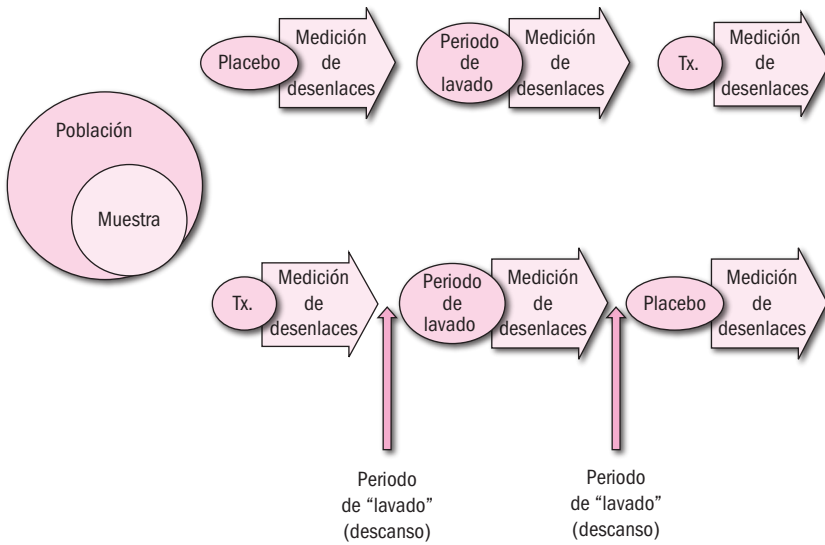


**Figura 5-4.** Diseños cruzados.

## Conducción del ensayo clínico

### Seguimiento y apego al protocolo de estudio

Un seguimiento apropiado es tan importante como un buen diseño dentro del ensayo clínico, ya que la diferencia del efecto está dada en buena medida por el poder de la muestra, por lo que las pérdidas o un pobre apego a la intervención pueden dejar un estudio con un bajo poder o resultados sesgados. De allí la importancia de contar con estrategias para garantizar un buen seguimiento y apego.

En este sentido existen estrategias conocidas para evitar las pérdidas y mantener un buen apego.<sup>14</sup> En el cuadro 5-3 se describen algunos principios y ejemplos que facilitan el seguimiento y apego de los pacientes en un ensayo clínico.

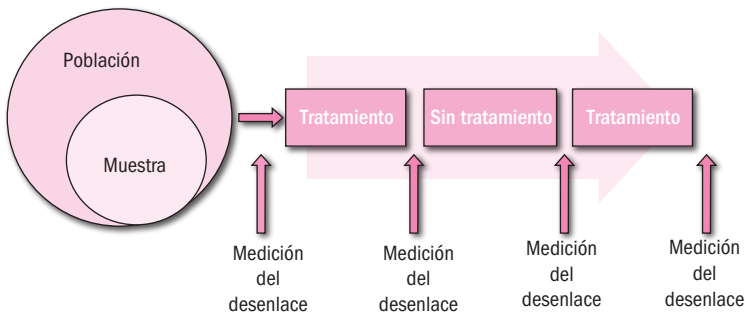


**Figura 5-5.** Series de tiempo.

**Cuadro 5-3.** Principios y estrategias para el seguimiento de los participantes en un ensayo clínico y el apego a la intervención

| Principio                                                      | Estrategias y ejemplos                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Elección apropiada de participantes                            | <p>Selección de personas que por sus características es más factible que terminen el estudio</p> <p>Para medir el apego pueden citarse a 2 o 3 consultas previas a la aleatorización</p> <p>Evitar sujetos que van a mudarse a otra región, aquellos con dificultades para la comprensión o comunicación. Pacientes que con cierta regularidad no acuden a citas programadas</p>                                                                                       |
| Intervenciones sencillas                                       | <p>Una sola intervención es siempre mejor</p> <p>Dosificaciones únicas y sencillas facilitan el apego</p>                                                                                                                                                                                                                                                                                                                                                              |
| Facilitar las visitas de estudio                               | <p>Las visitas hay que espaciarlas tanto como el estudio lo permita para que no se pierda el contacto con los participantes pero no se caiga en una rutina tediosa</p> <p>Es importante considerar aspectos como la coincidencia con otras citas al hospital, el reembolso de la transportación para esa consulta, evitar esperas en la atención y facilitar una buena relación de los participantes con el equipo</p>                                                 |
| Mediciones útiles, sencillas y no dolorosas                    | <p>Siempre que sea posible hay que seleccionar mediciones que sean fáciles de hacer, no dolorosas y que se puedan efectuar rápidamente. La pertinencia de éstas es invaluable</p> <p>Es importante dar a los participantes los resultados que se obtengan de dichas mediciones siempre que esto sea posible, además de contar con estrategias de referencia y manejo en caso necesario</p>                                                                             |
| Estrategias para que los participantes continúen en el estudio | <p>Mantener una relación cordial con cada uno de ellos, reiterarles la importancia de su participación y el valor científico del estudio son factores clave para que los participantes permanezcan durante el seguimiento. También es importante reiterarles que la presencia de eventos adversos no los excluye de su participación</p> <p>Recordatorios, tarjetas de cumpleaños o de Navidad pueden ser incentivos que faciliten la participación de los sujetos</p> |
| Búsqueda de los pacientes perdidos al seguimiento              | <p>Todos los pacientes incluidos en el estudio deben ser vigilados. Aquellos que por alguna situación dejen de acudir al seguimiento deben ser contactados a través de diferentes mecanismos (p. ej., familiares, teléfono, telegrama, correo electrónico, etc.)</p> <p>Los pacientes no se pierden por razones triviales y es necesario conocer estas razones</p>                                                                                                     |

## Monitoreo

El objetivo del monitoreo es asegurarse de que los participantes no están siendo expuestos a una intervención dañina, o bien se les está privando de un beneficio sustantivo. También es útil para tomar decisiones con respecto a la continuación del protocolo, particularmente si la pregunta de investigación no será factible de ser contestada.

Cuando las complicaciones asociadas a la intervención son mayores a los beneficios potenciales, el protocolo tiene que pararse. De manera similar, si la eficacia de la inter-

vención supera los estimados iniciales, éste debe pararse y ofrecer el beneficio observado a todos los participantes, ya que de lo contrario el proyecto no sería ético.

El monitoreo también es útil para valorar si la pregunta de investigación tiene bajas posibilidades de ser contestada o es necesario hacer ajustes. Un pobre apego a la intervención o una mala tasa de inclusión también son motivos de terminación temprana.

La terminación temprana de un ensayo clínico es una decisión compleja, ya que tiene que ponerse en la balanza la responsabilidad ética y el avance del conocimiento. Cuando se termina el protocolo antes de lo planeado, las probabilidades de obtener conclusiones sólidas disminuyen de manera considerable.<sup>14</sup>

## Análisis de los ensayos clínicos

### Punto fundamental

*La exclusión de los participantes aleatorizados o los desenlaces observados en el análisis, así como el análisis por subgrupos, pueden llevar a resultados sesgados de magnitud y dirección desconocidas.*

“Un ensayo clínico apropiadamente planeado y ejecutado es una técnica experimental poderosa para estimar la efectividad de una intervención.”<sup>16</sup> Este concepto ha sido aplicado en numerosos ensayos clínicos bajo la premisa de que todo ensayo clínico controlado comienza con la planeación cuidadosa del mismo, pasando por un proceso detallado de ejecución y monitoreo, siempre vigilando los procedimientos que garanticen la comparabilidad de los resultados.<sup>7</sup>

El análisis estadístico de las hipótesis primarias en un ensayo clínico es generalmente obvio; sin embargo, es imprescindible mencionar que la piedra angular de su análisis está fundamentada en un planteamiento meticuloso del diseño. Por otro lado, la estimación de los resultados se realiza a través de técnicas estadísticas propias a este diseño.<sup>7,14</sup>

Lo más común dentro de los ensayos clínicos es que se mida la incidencia de algún evento en los grupos de individuos seguidos en un determinado lapso y que este evento se exprese de manera dicotómica (por ejemplo, infarto agudo al miocardio, recurrencia del tumor, etc.), y la proporción de individuos que llegan a presentarlo.<sup>17</sup>

### Principio analítico por intención de tratar

Todos los ensayos clínicos aleatorizados deben ser analizados por “intención de tratar”; esto se refiere a que cada sujeto es analizado de acuerdo con la asignación original del tratamiento y sus efectos. Este tipo de análisis previene sobrestimar los resultados por la aparición de sesgos; estos últimos pueden originarse por la exclusión de sujetos después de la aleatorización; causas comunes son: sujetos que no reciben el tratamiento originalmente asignado, individuos que no reciben tratamiento, falta de apego al mismo o sujetos que mueran antes de que el tratamiento sea dado. El análisis por “intención de tratar” puede subestimar el efecto real del tratamiento, pero disminuye las posibilidades de obtener resultados sesgados.

## Análisis por protocolo

El análisis “por protocolo” es una alternativa al análisis por “intención de tratar”, ya que sólo analiza a un subgrupo de individuos que cumplieron suficientemente con el protocolo, lo que contrasta con el análisis más conservador de “intención de tratar”. En general se especifica durante la planeación el cumplimiento mínimo de la exposición al régimen de tratamiento, disponibilidad de mediciones de la variable primaria, elegibilidad correcta y ausencia de cualquier otra violación mayor al protocolo (errores en la asignación del tratamiento, uso de medicamentos “no permitidos”, pobre cumplimiento, datos faltantes, etc.). En este análisis también se excluyen los eventos que ocurrieron después de que el sujeto dejó de apegarse al protocolo, lo que con facilidad introduce sesgos, ya que es difícil determinar si la disminución en el apego está relacionada con el tratamiento (por ejemplo, es más tóxico y los participantes lo abandonan) y, por ende, con el resultado. Una de las grandes críticas a este tipo de análisis es que los sujetos que tienen un buen apego al estudio y al tratamiento suelen ser diferentes a aquellos que lo abandonan.<sup>7,14</sup>

Considerando las ventajas y desventajas potenciales de estos dos tipos de análisis, lo más conveniente es analizar los resultados tanto por “intención de tratar” como “por protocolo”.

En la literatura existen algunos ejemplos al respecto que vale la pena comentar y que ejemplifican las fortalezas y debilidades de estos análisis. En el estudio PEPI, por sus siglas en inglés (*Post-menopausal Estrogen-Progestin Interventions Trial*), se asignaron de manera aleatoria 875 mujeres posmenopáusicas a cuatro esquemas de reemplazo hormonal; sólo con estrógenos, estrógenos + progestágenos o placebo. Al cabo de tres años, 30% de las mujeres asignadas al grupo de manejo con estrógenos había abandonado el tratamiento por la presencia de hiperplasia endometrial, precursora del cáncer de endometrio. Si estos resultados sólo se hubieran analizado “por protocolo”, la asociación del tratamiento con estrógenos y cáncer endometrial se hubiera perdido.<sup>18</sup>

## Análisis por subgrupos

Este análisis tiene una reputación mixta, ya que se efectúan comparaciones de los participantes aleatorizados en uno o más subgrupos de la cohorte estudiada. Lo anterior puede prestarse fácilmente a un mal uso y puede conducir a conclusiones equivocadas, ya que los grupos suelen ser pequeños, en comparación con la muestra completa.

Con el cuidado debido, y preestableciendo las variables que pueden influir sobre el desenlace desde la planeación, este análisis puede ser una herramienta para mejorar la precisión y compensar por cualquier falla en el balance entre los grupos de tratamiento. Por tal motivo, y con el objeto de preservar la aleatorización los subgrupos, deben definirse por mediciones efectuadas previas a esta maniobra.<sup>7,14</sup>

Así, por ejemplo, en un estudio sobre uso de alendronato para la prevención de fracturas por osteoporosis se observó que el fármaco disminuía en 14% el riesgo de fractura en las mujeres con densidad ósea baja. El análisis por subgrupos reveló que el alendronato era efectivo (reducción del riesgo de fractura del 36%,  $p < 0.01$ ) en las mujeres cuya

densidad ósea se encontraba más de 2.5 desviaciones estándar por debajo de lo normal. En cambio, el tratamiento no fue eficaz en las mujeres con una densidad ósea mayor.<sup>6</sup> Es importante señalar que la aleatorización se preservó, ya que se comparó la frecuencia de fracturas en el grupo de alendronato con la frecuencia de fracturas en el grupo control para cada uno de los subgrupos.

## Interpretación de los resultados. Importancia y validez de los resultados del ensayo

La interpretación y aplicación correcta de los resultados de un ensayo clínico son tan importantes como el análisis mismo. Las pruebas estadísticas utilizadas para el análisis del ensayo clínico no son diferentes a las otras empleadas en los distintos diseños, y la selección de éstas depende de la pregunta y tipo de variable. Así pues, la chi cuadrada, el análisis de supervivencia o la *t* de Student pueden decir si existen o no diferencias; sin embargo, no explican si el efecto del tratamiento es lo suficientemente importante (magnitud en el efecto del tratamiento) para que éste sea útil en los pacientes. Es importante señalar que los estudios que reclutan gran cantidad de participantes pueden encontrar diferencias “triviales” desde el punto de vista estadístico, por lo que es fundamental reconocer las diferencias mínimamente importantes y que son significativas a los pacientes, al clínico, proveedores, etcétera; para ello se necesita un proceso que involucra las matemáticas y el juicio clínico.

**Determinación de las diferencias importantes.** En el ensayo clínico de dos grupos paralelos en donde se evalúa la incidencia de un evento (por ejemplo, infarto agudo al miocardio) se determina la frecuencia de eventos (puede ser a través también del riesgo relativo dado que se evalúa la incidencia) en el grupo control (FEC) y en el grupo experimental (FEE). Cuando se tiene que ajustar por la variable tiempo, la FEC y la FEE se estiman de acuerdo con la probabilidad de ocurrencia del evento (por ejemplo, falla) obtenidas de la curva de Kaplan y Meier (véase cap. 13).

Así, por ejemplo, en un ensayo donde se probaron dosis altas (grupo experimental) vs. dosis bajas (grupo control) de ácido acetilsalicílico para la prevención de la enfermedad cerebrovascular, infarto agudo al miocardio o muerte un mes después de la endarterectomía carotídea, la FEE en el grupo de altas dosis fue de 8.2% y de 3.7% en el grupo control ( $p = 0.002$ ).<sup>19-20</sup>

Desde el punto de vista matemático, lo anterior resulta sencillo y es fácil de interpretar para el clínico; sin embargo, es muy importante decidir si estas diferencias son relevantes. Es en esta segunda etapa en donde el buen juicio juega un papel trascendental y puede dar más y mejor información a los clínicos y a los pacientes.

**Las diferencias mínimamente importantes para el clínico.** En general éstas se expresan a través del número de pacientes que es necesario tratar con la terapia experimental para prevenir un evento adverso (por ejemplo, progresión de la enfermedad) o que causen daño por el mismo. Para ello es necesario calcular la reducción del riesgo absoluto (RRA), expresado como un porcentaje que se obtiene al sustraer la proporción

de individuos que presentaron el evento de interés en el grupo experimental (FEE) de la proporción de individuos que lo hicieron en el grupo control (FEC). Esta cifra demuestra en forma de porcentaje la reducción del riesgo de un evento (muerte, recurrencia, etc.) en relación con lo que ocurre en los pacientes del grupo control. La ventaja de esta medida de impacto es que preserva el riesgo basal de la ocurrencia del evento, independientemente de la terapia empleada.

Una desventaja del RRA es que puede ser difícil de interpretar, y puede no tener un significado claro para los pacientes, ya que la utilidad de una terapia está dada no sólo en función de la reducción relativa del riesgo, sino también del riesgo del desenlace adverso que se quiere prevenir (por ejemplo, la muerte). Asimismo, es útil conocer si el beneficio ofrecido por la nueva terapia retribuye el esfuerzo y costo en su adquisición o implantación. Esto último justamente es conocido como el número necesario de pacientes a tratar (NNT), que informa de una manera clara y sencilla cuál es el esfuerzo que como clínicos y pacientes deben poner para prevenir un evento. En el sentido opuesto se puede calcular el número necesario a dañar (NND) a partir de la frecuencia de eventos adversos del grupo experimental y del grupo control.

El número necesario de pacientes a tratar se obtiene al calcular la inversa de la reducción del riesgo absoluto ( $NNT = 1/RRA$ ). Por ejemplo, si en una población de enfermos, en un determinado periodo el riesgo de muerte fuera de 1% y la nueva terapia disminuyera el riesgo de fallecer en 25%, la RRA sería de 0.0025 (o 25 muertes en 10 000 pacientes tratados). El NNT correspondiente ( $NNT = 1/0.0025 = 400$ ) sería de 400 pacientes durante un tiempo determinado. En el cuadro 5-4 se detallan algunos ejemplos de estas medidas de riesgo empleadas para la interpretación de los ensayos clínicos.<sup>17,19</sup>

**Cuadro 5-4.** Utilidad del NNT en diversos estudios clínicos

| Evento                                                                                | Intervención                                | FEC (%) | FEE (%) | Diferencia de riesgo | Tiempo de seguimiento | NNT |
|---------------------------------------------------------------------------------------|---------------------------------------------|---------|---------|----------------------|-----------------------|-----|
| Pacientes $\geq 60$ años con IAM y fallecimiento $>2$ años (a)                        | Beta-bloqueadores por 1 mes                 | 9.8     | 7.3     | 2.5                  | 2 años                | 40  |
| Pacientes $\geq 60$ años con IAM y fallecimiento a las 5 semanas (b)                  | Estreptocinasa                              | 12.0    | 9.2     | 2.8                  | 5 semanas             | 36  |
| Hiperplasia prostática benigna y prevención de cirugía (c)                            | Finasterida                                 | 10      | 4.6     | 5.4                  | 4 años                | 18  |
| Estenosis carotídea grave sintomática y enfermedad cerebrovascular grave o muerte (d) | Endarterectomía versus mejor terapia médica | 18      | 8       | 10                   | 2 años                | 10  |

a) JAMA 1982;247:1707-14

b) Lancet 1988;2:349-60

c) New Engl J Med 1998;338:557-63.

d) New Engl J Med 1991;325:445-53.



## Aspectos éticos de la investigación

Proteger los derechos y el bienestar de los individuos que participan en investigaciones médicas es un deber ético de los investigadores y de la sociedad misma. Realizar un ensayo clínico puede generar inquietudes de orden ético y sólo debe realizarse si cumple con la premisa de que al momento del inicio no existen evidencias de que uno de los tratamientos (experimental o control) sea superior a otro.<sup>17,21</sup>

El diseño experimental significa que los tratamientos son asignados por un investigador, no por su médico tratante ni son elegidos por el propio paciente. Además, cuando la asignación es mediante un sorteo o cuando se efectúa algún tipo de enmascaramiento, los aspectos éticos se vuelven aún más complejos, y éstos quedan por encima de los aspectos científicos.<sup>7,17</sup> Por otro lado, y como en cualquier investigación médica, es fundamental salvaguardar la confidencialidad de la información e intimidad de los participantes, por lo que se deben adoptar las medidas necesarias para proteger la intimidad de los sujetos durante la recolección de la información (sólo obtener datos íntimos si son imprescindibles, almacenar los cuestionarios en lugar protegido, etc.), automatización de los datos (disociación de información y encriptado de ficheros, claves de acceso a los ficheros, destrucción de archivos cuando ya no son necesarios, etc.) y publicación de resultados (no publicar fotografías u otros datos que permitan conocer la identidad de los individuos, etc.).<sup>22</sup>

Los principios éticos de la investigación están enmarcados en las recomendaciones para realizar investigación clínica en seres humanos. La primera de estas recomendaciones fue el Código de Nüremberg de 1947, seguido de los códigos de la Asociación Médica Mundial, la Declaración de Ginebra de 1948 y la Declaración de Helsinki de 1964, con sus actualizaciones posteriores (Corea, 2008) (véase capítulo 7).

En 1974, el informe Belmont de la *National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research* de Estados Unidos, resumió en tres los principios éticos de cualquier investigación médica: *autonomía*, *beneficencia* y *justicia*. Más tarde, se añadió el principio de **no maleficencia**, generalizando la aplicación de estos principios a la medicina clínica.

Estos cuatro principios buscan la protección de los individuos y enfatizan la importancia de que los participantes en estudios médicos tomen decisiones voluntarias e informadas sobre las intervenciones a las que puedan ser sometidos (*principio de autonomía*); de allí la obligación de que los investigadores obtengan el consentimiento informado sin coerción, intimidación o incentivos indebidos.<sup>7,22</sup>

En éstos, también se menciona de manera explícita la prioridad de no cometer daño a los individuos (*no maleficencia*) y de maximizar el beneficio para los participantes (*beneficencia*), obligando a que las comparaciones en los estudios experimentales se realicen utilizando las mejores terapias disponibles al momento de la investigación. Compromete a los investigadores a dar a conocer a los participantes cualquier información relevante derivada del estudio.

El principio de *justicia* está encaminado a la justicia distributiva, y establece que los riesgos y beneficios derivados de las investigaciones se repartan de forma razonable en

la sociedad; esto abarca la protección de los individuos más vulnerables. En torno a este principio existe un debate que cuestiona realizar investigaciones en países menos desarrollados utilizando criterios diferentes a los requeridos en países ricos.<sup>22</sup>

Aunque estos dilemas éticos pueden llegar a ser de difícil solución, el consentimiento informado y los comités institucionales de investigación en humanos son importantes salvaguardas de la ética, y en la actualidad la investigación médica sería impensable sin éstos.

## Comités de Ética y Conferencia Internacional de Armonización y Buenas Prácticas Clínicas

En toda institución donde se realiza investigación médica es necesario que exista un Comité Institucional (de Investigación y Ética) con capacidad para asegurar: 1) que la investigación propuesta responda a las necesidades de salud de la población; 2) que no se exponga a los participantes a riesgos inaceptables e innecesarios, y 3) que los participantes potenciales tengan la garantía de ser completamente informados y, por tanto, puedan decidir sobre su participación tomando en cuenta las consecuencias previstas.<sup>7</sup>

El Comité de Ética está formado por un grupo de personas con experiencia en investigación y diversos campos de la medicina, además de que es indispensable que haya representantes de la sociedad, de los pacientes y expertos en bioética. Sus miembros promueven, documentan, vigilan y evalúan la investigación en seres humanos y se encargan de revisar en forma periódica el avance de los diferentes proyectos, los eventos adversos y asegurarse de que no se ponga en riesgo a ninguno de sus participantes en ningún momento (véase capítulo 7).

## Conferencia Internacional de Armonización y Buenas Prácticas Clínicas

Las buenas prácticas clínicas (BPC) constituyen un estándar para el diseño, conducción, desarrollo, monitoreo, auditoría, registro, análisis y reporte de estudios de investigación clínica en los que participan seres humanos como sujetos de estudio. Fueron creadas en 1995 por la Organización Mundial de la Salud y pretenden: 1) la protección de los derechos, la seguridad y el bienestar de los sujetos de estudio, y 2) la credibilidad de los datos generados, necesaria para el registro de una innovación terapéutica en cualquier parte del mundo. Su cumplimiento facilita la conducción y análisis, así como la protección de los participantes.<sup>23</sup>

## Fases de los diseños experimentales

El desarrollo de nuevos medicamentos, vacunas y dispositivos médicos tiene que pasar por una serie de fases antes de poder comercializarlos y usarlos en cualquier paciente. Para su mejor comprensión existe una nomenclatura común que describe el orden que deben seguir estos estudios desde las fases más incipientes hasta la aprobación y comercialización.<sup>2,24</sup>

**Fase preclínica.** Cuando existe una molécula prometedora, las primeras pruebas se realizan en animales de experimentación con el objeto de determinar: 1) la distribución del fármaco en diferentes tejidos y órganos; 2) toxicidad en diferentes células, tejidos y órganos (farmacodinamia); 3) metabolismo y vías de excreción, y 4) dosis letal al 50%. Si la dosis letal al 50% está francamente por arriba del efecto farmacodinámico deseado, si los otros atributos son favorables, y si parece ser un producto factible y económicamente rentable de desarrollar, entonces se inician los estudios en humanos.

**Fase I.** En esta fase se utilizan grupos pequeños de voluntarios sanos (de 20 a 80 personas); están diseñados para establecer los posibles efectos tóxicos, absorción, distribución y metabolismo en humanos. Las drogas con efectos tóxicos serios suelen probarse también en pacientes gravemente enfermos que no han respondido a las mejores terapias disponibles. Son estudios altamente monitoreados y se cuantifican de manera extensa los efectos farmacocinéticos y farmacológicos que permiten planear la siguiente fase.

**Fase II.** Esta fase se prueba en grupos de 50-100 sujetos enfermos relacionados con la enfermedad o condición. Su objetivo principal es establecer la seguridad y eficacia, por lo que están estrechamente controlados y monitoreados. Demostrar la efectividad es igualmente importante y es lo que permite decidir si se efectuará un ensayo de mayor tamaño.

**Fase III.** Los estudios en esta fase se realizan cuando existe evidencia preliminar que sugiere efectividad del compuesto de prueba. Son estudios a gran escala en pacientes con la enfermedad de interés. Son aleatorizados y su objetivo es probar y comparar el nuevo fármaco con pacientes manejados con el tratamiento estándar.

**Fase IV.** Esta fase corresponde a la última de este proceso y se da una vez que la droga ha sido aprobada, por lo que es necesario establecer algún tipo de vigilancia. El objetivo de estos estudios es obtener información adicional sobre la seguridad y eficacia del nuevo medicamento, conocer las interacciones con otros medicamentos y alimentos, la distribución y determinantes de su uso (farmacoepidemiología) y su costo-efectividad.<sup>2,5,7</sup>

---

## Referencias

1. Piantadosi S. *Clinical trials. A methodologic perspective*. Wiley Series in Probability and Statistics. (Chapter 2) Nueva York, EUA. pp. 7-28. 1997.
2. Friedman LW, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. (Chapter 1) Springer-Verlag. Nueva York, EUA. pp. 1-15. 1998.
3. Agency for Healthcare Research and Quality. (Acceso en diciembre 23, 2009, en <http://www.ahrq.gov/>).
4. Cummings SR, Grady DG, Hulley SB. Designing a randomized blinded trial. En: Hulley SB, Cummings RB, Browner WS, Grady DG, Newman T. *Designing clinical research*. (Chapter 10) 3a. ed. Lippincott Williams & Wilkins. Filadelfia, EUA. pp. 147-161. 2007.
5. Sackett D. The tactics of performing therapeutic trials. En: *Clinical Epidemiology. How to do clinical practice research*. (Chapter 5) 3a. ed. Lippincott Williams & Wilkins. Filadelfia, EUA. pp. 66-172. 2006.

6. Cummings SR, Black DM, Thompson DE, Applegate WB, Barrett-Connor E, Musliner TA *et al.* Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the fracture intervention trial. *JAMA* 280:2077-2082. 1998.
7. Lazcano-Ponce E, Salazar-Martínez E, Gutiérrez Castrellón P, Ángeles Llerenas A, Hernández Garduño A, Viramontes JL. *Ensayos clínicos aleatorizados: variantes, métodos de aleatorización, análisis, consideraciones éticas y regulación.* *Salud Pub Mex* 46:559-584. 2004.
8. Schafer A. The ethics of the randomized controlled trials. *N Engl J Med* 307:719-724. 1982.
9. Freedman B. Placebo-controlled trials and the logic of clinical purpose. *IRB* 12:1-6. 1990.
10. Kienle GS, Kiene H. The powerful placebo effect: Fact or fiction? *J Clin Epidemiol* 50:1311-1318. 1997.
11. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 331:394-398. 1994.
12. Lurie P, Wolfe SM. Unethical trials of intervention to reduce perinatal transmission of the human immunodeficiency virus in developing countries. *N Engl J Med* 337:853-856. 1997.
13. Schulz KF, Grimes DA. Blinding in randomized trials. Hiding who got what. *Lancet* 359: 696-700. 2002.
14. Grady D, Cummings SR, Hulley SB. Alternative trial designs and implementation issues. En: Hulley SB, Cummings RB, Browner WS, Grady DG, Newman T. *Designing clinical research.* 3a. ed. (Chapter 11) Lippincott Williams & Wilkins. Filadelfia, EUA, pp. 163-182. 2007.
15. Ridker PM, Cook NR, Lee I, Gordon D, Gaziano JM, Manson JE *et al.* A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med* 352:1293-1304. 2005.
16. Friedman LW, Furberg CD, DeMets DL. *Fundamentals of clinical trials.* (Chapter 6) Springer-Verlag New York. Nueva York, pp. 82-93. 1998.
17. Calva-Mercado, JJ. *Estudios clínicos experimentales.* *Salud Pub Mex* 42:349-358. 2000.
18. Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. *JAMA* 273:199-208. 1995.
19. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. How to practice and teach EBM. (Chapter 5), 2a. ed. Churchill Livingstone, Londres, RU, pp. 105-153. 2000.
20. Taylor DW, Barnett HJM, Haynes RB, Ferguson G, Sackett D, Thorpe K. ASA and carotid endarterectomy (ACE) Trial Collaborators. Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomized controlled trial. *Lancet* 353: 2179-2184. 1999.
21. Peto R, Baigent C. Trials: the next 50 years. Large scale randomized evidence of moderate benefits. *Br Med J* 317:1170-1171. 1998.
22. Verástegui E. Consenting of the vulnerable: The informed consent procedure in advanced cancer patients in Mexico. *BMC Medical Ethics* 7:1. 2006.
23. Guidelines for good clinical practice (GCP) for trials on pharmaceutical products. World Health Organization. WHO Technical Report Series, No. 850, 1995, Annex 3.
24. Sackett D. The principles behind the tactics of performing therapeutic trials. En: *Clinical Epidemiology. How to do clinical practice research.* (Chapter 6) 3a. ed. Lippincott Williams & Wilkins. Filadelfia, EUA, pp. 173-243. 2006.



# Capítulo 6

## Farmacovigilancia

*María Eugenia Jiménez Corona  
Alma Rosa González Montiel  
Fátima del Carmen Aguilar Díaz*

### Introducción

Después de la Segunda Guerra Mundial la llamada **explosión farmacológica** permitió grandes avances en el tratamiento de enfermedades que antes conducían de manera rápida e inevitable a la muerte o a una grave incapacidad. Este progreso se ha acompañado de accidentes; uno de los más conocidos es el atribuido a la talidomida, asociada con casos de focomelia.

Desde entonces la preocupación por la seguridad de los fármacos ha sido un tema relevante en el desarrollo de medicamentos, biológicos (incluyendo las vacunas) y dispositivos médicos.

Con la finalidad de incrementar la seguridad para el paciente, se ha promovido el contar con métodos para la evaluación de los beneficios y los riesgos potenciales de cualquier intervención terapéutica. Al utilizar un medicamento para curar (antibióticos), atenuar (analgésicos), o diagnosticar una patología (p. ej., yodo-131), el usuario está expuesto al riesgo de sufrir una reacción no deseada. Todo medicamento tiene la capacidad de causar reacciones adversas; sin embargo, es importante destacar que estas reacciones varían desde algunas pequeñas molestias hasta efectos tan graves que ponen en peligro la vida del paciente.

Lo antes mencionado ha llevado a que desde hace algunas décadas los medicamentos antes de su comercialización sean sometidos a pruebas de seguridad mediante estudios toxicológicos en animales (fase preclínica) y su posterior evaluación en ensayos clínicos en humanos (fases I-III), para garantizar su eficacia y seguridad (véase cap. 5).

Sin embargo, en la etapa de poscomercialización, pueden aparecer reacciones adversas que no fueron identificadas en los estudios clínicos, por lo que es necesario continuar con su evaluación en dicha etapa poscomercialización (fase IV), es decir, una vez que el medicamento ya se encuentra disponible en el mercado (figura 6-1).

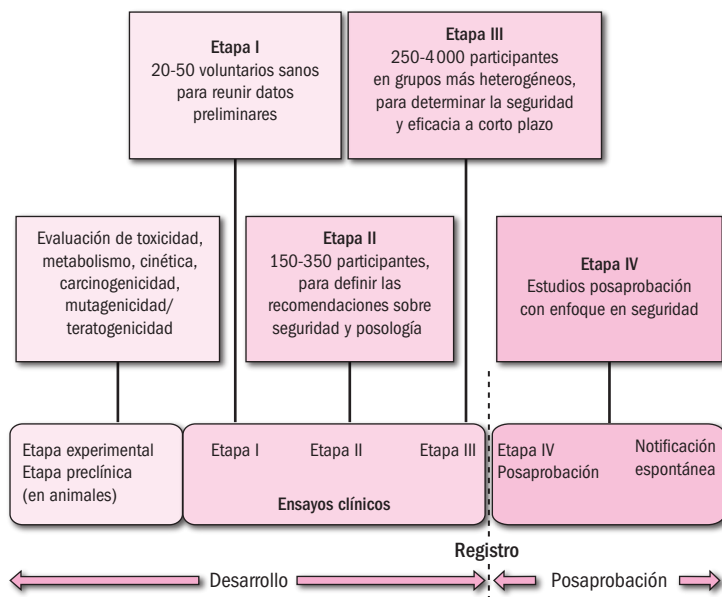


**Figura 6-1.** Etapas para el desarrollo de un medicamento.

## Antecedentes

En la historia de la Medicina existen reportes de varios desastres originados por el uso de medicamentos (cuadro 6-1). En 1847 se introdujo el cloroformo como anestésico y, en 1848, una niña de 15 años fue sometida a cirugía y anestesiada con cloroformo y murió debido a un episodio de fibrilación ventricular. A raíz de esto, la revista *Lancet* invitó a la comunidad médica de Gran Bretaña a reportar las muertes relacionadas con el uso del cloroformo como anestésico; los hallazgos fueron publicados en la misma revista en 1893; este hecho se considera el antecedente del sistema de reporte espontáneo para una “supuesta reacción adversa”.

En junio de 1906 fue aprobada por el congreso de EUA la *Ley original de alimentos y medicamentos*, la cual estableció que los medicamentos deberían cumplir con ciertos requisitos, tales como pureza y estar libres de contaminación; sin embargo, no se establecían requisitos de eficacia. Entre los decenios de 1930-1939 y 1940-1949, llamada la “era de la terapéutica farmacológica”, se introdujo el uso de la penicilina, considerando la posibilidad de que los medicamentos podían producir efectos adversos. En 1930, el nombre de la Administración de Alimentos, Medicamentos e Insecticidas se redujo a Federación de Alimentos y Medicamentos (*Food and Drug Administration*, FDA).

En 1937, en EUA, ocurrió el primer accidente grave: se comercializó un jarabe de sulfonilamida con dietilenglicol que dio lugar a más de 100 muertes. A raíz de este evento se emitió en Estados Unidos el Acta Federal de Alimentos, Drogas y Cosméticos, en 1938,

**Cuadro 6-1.** Eventos relevantes que dieron origen a la farmacovigilancia

| Año/lugar                                                                         | Sustancia                                                 | Accidente                                                                | Evento                                                                                  | Resultado                                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1847<br>Inglaterra                                                                | <b>Cloroformo</b><br>como anestésico                      | Muere un niño<br>de 15 años                                              | <i>Lancet</i> solicita<br>reporte de<br>muertes<br>similares                            | En 1893 se publican<br>resultados, antecesor del<br>REPORTE ESPONTÁNEO                                                                                                                                      |
| 1930-1940                                                                         | Era terapéutica<br>farmacológica.<br><b>Penicilinas</b>   | Se observan<br>algunos efectos<br>adversos.<br>Primer accidente<br>grave | Se concluye<br>que los<br>medicamentos<br>pueden causar<br>eventos adversos             | Se inicia la modificación<br>legislativa                                                                                                                                                                    |
| 1937<br>Estados<br>Unidos                                                         | Jarabe con<br><b>dietilenglicol y<br/>sulfanilamidas</b>  | Más de 100<br>muertes<br>reportadas                                      | Se emiten<br>normas<br>destinadas a<br>verificar la<br>seguridad de los<br>medicamentos | <b>Se crea la Food and Drug<br/>Administration, FDA,<br/>primera agencia<br/>regulatoria</b>                                                                                                                |
| 1960/<br>Alemania<br>y otros<br>países<br>1964/<br>Reino unido<br>1968/<br>Europa | Revolución<br>de los<br>medicamentos<br><b>Talidomida</b> | Epidemia<br>de focomelia                                                 | Se asocia a la<br>Talidomida                                                            | Europa restablece el<br>reporte espontáneo<br>de farmacovigilancia.<br>Se introduce el uso de<br>la tarjeta amarilla y la<br>legislación para regular<br>los medicamentos.<br>Establecen directiva<br>65/65 |

que fue la primera agencia reguladora de medicamentos en el mundo, la cual estableció un nuevo sistema de regulación de fármacos.

A principios de la década de 1960-1969, en Europa existía la llamada “revolución de los medicamentos”, época en la que el uso de éstos se encontraba en pleno apogeo y se tenía amplia confianza en su uso para combatir las enfermedades. En Alemania, Gran Bretaña y Australia se presentó una epidemia de focomelia en hijos de madres que habían ingerido talidomida durante el embarazo, que se caracterizaba por una aplasia de los huesos largos de las extremidades, por lo que manos y pies provenían directamente del torso. Tal desastre dio una nueva percepción de los posibles riesgos por el uso de medicamentos. Inicialmente se pensó en factores hereditarios, pero su carácter epidémico indujo a considerar la intervención de factores externos, tales como infecciones virales, radiaciones o alimentos.

En 1961, W. Lenz, en una reunión de la Sociedad de Pediatría, sugirió la asociación de la malformación con el uso de la talidomida durante el embarazo, se realizó un estudio de casos y controles, y en 1961 se retiró del mercado el medicamento en Alemania y posteriormente en otros países. El evento histórico con la talidomida estimuló el desarrollo de sistemas de farmacovigilancia de reporte espontáneo; así nació en Europa, en el



Reino Unido, en 1964, el sistema de reporte mediante la “Tarjeta amarilla”; asimismo, se implementaron cambios en la legislación para regular los medicamentos. En 1971, el Comité Dunlop, el predecesor del Comité de Seguridad de los Medicamentos del Reino Unido, estableció lineamientos sobre la seguridad de los medicamentos.

Un ejemplo más reciente, sobre la utilidad de la farmacovigilancia, es el caso del medicamento cerivastatina, aprobado como agente regulador de los lípidos en 1997; a finales del año 2000 se habían notificado al Centro Colaborador de la Organización Mundial de la Salud (OMS) para la Vigilancia Farmacéutica Internacional de Uppsala, un total de 549 casos de rabdomiólisis. En junio de 2001 se adoptaron medidas reglamentarias de alcance paneuropeo para declarar contraindicado el uso simultáneo de cerivastatina y gemfibrozilo. El 8 de agosto de 2001 el fabricante retiró voluntariamente la cerivastatina del mercado aduciendo que incrementaba el riesgo de rabdomiólisis, sobre todo combinada con gemfibrozilo.

## Definiciones

**Farmacovigilancia.** De acuerdo con la OMS (2002), la farmacovigilancia (o vigilancia de medicamentos) es la ciencia que trata de recoger, vigilar, investigar y evaluar la información sobre los efectos de los medicamentos, productos biológicos, plantas medicinales y medicinas tradicionales, con el objetivo de identificar información de nuevas reacciones adversas, conocer su frecuencia y prevenir los daños en los pacientes.

**Reacción adversa a medicamentos.** Para fines operativos la Organización Mundial de la Salud (OMS) define la **reacción adversa a los medicamentos** (RAM) como una respuesta nociva e indeseable que ocurre a dosis normalmente usadas en el ser humano para la profilaxis, el diagnóstico o la terapia de una enfermedad, o para la modificación de una función fisiológica, pudiendo tratarse de una reacción esperada cuando ésta se conoce y está contemplada en la literatura científica. De igual forma, puede tratarse de una reacción inesperada cuando su naturaleza o gravedad no está descrita en la literatura científica, ni en la información contenida en la etiqueta o en la información para prescribir, ni en la documentación presentada para su registro sanitario, además de que no es posible inferirla de su actividad farmacológica.

**Evento adverso.** Cualquier ocurrencia médica “adversa” que puede aparecer durante el tratamiento con un producto farmacéutico, pero que no necesariamente tiene una relación causal con el tratamiento.

## Farmacovigilancia a nivel internacional

El desastre de la talidomida tuvo consecuencias que influyeron de manera positiva en la relevancia que cobró la seguridad del paciente; favoreció la definición de estrategias para incrementar la seguridad en los medicamentos; así, se destacaron las siguientes acciones: 1) los gobiernos empezaron a exigir a las compañías farmacéuticas pruebas de toxicidad



en animales más exhaustivas; 2) los ensayos clínicos controlados se promovieron como una herramienta básica para que los nuevos medicamentos demostraran su eficacia y seguridad antes de su autorización para ser comercializados, y 3) se propusieron diversas estrategias para evitar accidentes similares al ocurrido con la talidomida, que tomaron forma en lo que hoy se conoce como **farmacovigilancia**.

La historia de la farmacovigilancia internacional comenzó hace más de 50 años, cuando la vigésima Asamblea Mundial de la Salud (1959) acordó una resolución para iniciar un proyecto de viabilidad de un sistema internacional de seguimiento de las reacciones adversas a los medicamentos. Esta resolución fue la base del **Programa Internacional de Farmacovigilancia de la OMS**, el cual inició como un programa de **vigilancia farmacéutica internacional** en 1968, con la idea de compilar los datos existentes sobre las reacciones adversas a los medicamentos; era un proyecto piloto en donde participaban 10 países que contaban con un sistema de notificación de reacciones adversas. Actualmente, más de 100 países tienen sistemas nacionales de notificación de reacciones adversas a medicamentos (RAM) que reportan a la base de datos del Centro Colaborador de la OMS para la vigilancia farmacéutica internacional, denominado **Uppsala Monitoring Centre (UMC)** ubicado en Uppsala, Suecia. Algunas de las actividades más relevantes de este centro son:

- Recolectar y analizar reportes de RAM a nivel mundial.
- Comunicar problemas relacionados con la seguridad de los medicamentos.
- Apoyar activamente y dar formación en farmacovigilancia.
- Desarrollar la ciencia de la farmacovigilancia.

Cuando aparecen señales de problemas relacionados con la seguridad de los medicamentos, la OMS comparte los resultados con todos los países miembros, con la finalidad de tomar oportunamente medidas al respecto. Para llevar a cabo una adecuada farmacovigilancia se requiere de la participación de personal capacitado en diferentes disciplinas: médicos, epidemiólogos, farmacólogos clínicos, enfermeras y químicos, entre otros.

La farmacovigilancia es necesaria en cada país, ya que pueden existir diferencias entre países en la manifestación de reacciones adversas a medicamentos, vacunas y otros insumos para la salud y problemas relacionados con los mismos. Todo esto puede deberse a diferencias en la producción, distribución, calidad y composición de los mismos; incluso las reacciones pueden ser debidas a diferencias étnicas y raciales, entre otros factores. Asimismo, la farmacovigilancia es necesaria para la prevención de riesgos de los medicamentos en los seres humanos y evitar los costos económicos asociados a los efectos adversos no esperados.

La vigilancia sobre los medicamentos durante su comercialización y su utilización por un número considerablemente mayor de personas de las que participan en los ensayos clínicos y por grupos de población diferentes (p. ej., grupos de riesgo como niños y ancianos, poblaciones sometidas a condiciones diferentes de calidad de vida, etc.), son un componente importante de las acciones a cargo de los organismos reguladores de los medicamentos.

La OMS y los organismos sanitarios relacionados con los medicamentos, las agencias regulatorias nacionales, tales como la FDA en Estados Unidos; la Agencia Europea de Medicinas (*European Medicines Agency*, EMA) en Europa; y la Comisión Federal de Protección contra Riesgos Sanitarios (COFEPRIS) en México, entre otras, se han encargado de organizar sistemas que facilitan la detección de las reacciones adversas provocadas por los medicamentos, con el fin de limitar en lo posible los riesgos en las personas que los utilizan. Los gobiernos tienen la responsabilidad de garantizar la calidad, seguridad y eficacia de los medicamentos, vacunas y dispositivos médicos que se comercializan en cada país y de establecer normas para reglamentar la utilización de los mismos a favor de la defensa de la salud de los consumidores.

## ¿Cómo funciona un sistema de farmacovigilancia?

La farmacovigilancia es una actividad de responsabilidad compartida entre todos los agentes que utilizan un medicamento o vacuna, como son: autoridad sanitaria reguladora, laboratorios productores, titular del registro, centros de investigación, profesionales de la salud y pacientes. La información sobre las reacciones adversas puede generarse por notificación voluntaria de los médicos en el ejercicio privado o público en la consulta externa u hospitalaria, en centros previamente designados o por aplicación de técnicas epidemiológicas que permitan obtener información sistemática de diversas fuentes, por los propietarios de los registros de los productos, por parte de los laboratorios productores, así como por los propios pacientes.

Un sistema de farmacovigilancia “ideal” debe reunir ciertas características o atributos, como los que se describen en el recuadro de la siguiente página (figura 6-2).

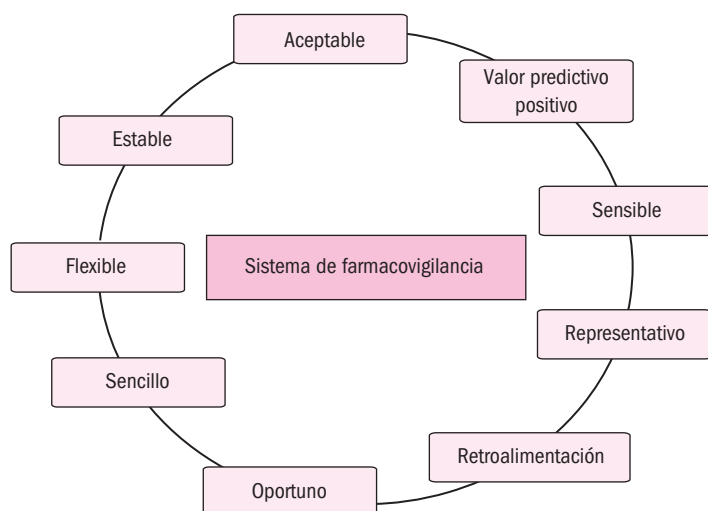


**Figura 6-2.** Características de un sistema de farmacovigilancia.

- **Sencillez.** Debe ser sencillo pero permitir el cumplimiento del objetivo.
- **Flexible.** Es importante que permita el registro de diferentes manifestaciones clínicas que se presentan en diferentes productos.
- **Aceptable.** Por todos los participantes involucrados en el sistema.
- **Sensible.** Debe valorar la capacidad de identificar la mayor proporción posible de casos que ocurre en la población.
- **Valor predictivo positivo.** Tiene que permitir que se verifique el número de casos reportados que en realidad sean casos.
- **Representativo.** Es necesario que realmente represente la ocurrencia del problema que se vigila en el tiempo.
- **Oportuno.** Se valorará el tiempo en el que ocurre el evento comparado con el tiempo en que se reporta.
- **Estable.** Fiabilidad de los reportes evaluando lo que se reporta en su sistema de vigilancia y la información que contiene la OMS.
- **Retroalimentación.** Hacia el Sector Salud y la industria farmacéutica, con la consecuente aplicación de esta información para la toma de decisiones, tanto para el Sector Salud como para los productores.

Los sistemas de farmacovigilancia recolectan, registran y evalúan de forma sistemática la información respecto a reacciones adversas de los medicamentos cuando son usados en la etapa poscomercialización por una población en condiciones naturales. Las estrategias o métodos de mayor utilización para recolectar información sobre los eventos adversos se describen a continuación (véase cuadro 6-2):

- Sistema de notificación espontánea: sistema de ámbito regional o nacional para notificar sospechas de reacciones adversas; es el método primario en farmacovigilancia.

**Cuadro 6-2.** Estrategias o métodos de recolección de información sobre reacciones adversas a medicamentos (RAM)

| Estrategia/método                   | Tipo de información                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sistemas de notificación espontánea | <ul style="list-style-type: none"> <li>• Reporte espontáneo de reacciones adversas y/o publicaciones</li> </ul>                                                                                                                                                                                                                                                                                                                             |
| Sistemas de vigilancia intensiva    | <ul style="list-style-type: none"> <li>• Farmacovigilancia activa</li> <li>• Centros centinela</li> </ul>                                                                                                                                                                                                                                                                                                                                   |
| Estudios epidemiológicos            | <ul style="list-style-type: none"> <li>• Reporte de casos</li> <li>• Estudios de cohortes</li> <li>• Estudios de casos y controles</li> <li>• Ensayos clínicos controlados</li> <li>• Metaanálisis</li> <li>• Estudios de morbilidad y mortalidad</li> <li>• Fusión de registros clínicos</li> <li>• Monitorización intensiva de pacientes hospitalizados</li> <li>• Monitorización de acontecimientos ligados a la prescripción</li> </ul> |

- Sistemas de vigilancia intensiva.
- Estudios epidemiológicos.

Los programas de farmacovigilancia tienen los siguientes objetivos:

- Detección temprana de las reacciones adversas e interacciones desconocidas hasta ese momento.
- Detección de aumentos de la frecuencia de reacciones adversas (conocidas).
- Identificación de factores de riesgo y de los posibles mecanismos subyacentes de las reacciones adversas.
- Estimación de los aspectos cuantitativos de la relación riesgo/beneficio y difusión de la información necesaria para mejorar la regulación y prescripción de medicamentos.
- Contribuir a la evaluación de los beneficios, daños, la eficacia y los riesgos que puedan presentar los medicamentos, conduciendo a la prevención de los daños y maximización de los beneficios.
- Prevención de los posibles riesgos derivados del uso de medicamentos.

Sobre las reacciones inesperadas y perjudiciales por el uso de medicamentos, se puede mencionar lo siguiente sobre las reacciones adversas medicamentosas (RAM):

- Las RAM están entre las 10 principales causas de defunción en todo el mundo.
- La mayoría de las RAM son evitables.
- Las RAM afectan a personas de todos los países.
- En algunos casos, los costos asociados a las RAM, por ejemplo, en relación con la hospitalización, la cirugía y la pérdida de productividad, sobrepasan el costo de los medicamentos.
- No hay medicamentos exentos de riesgos. Una evaluación atenta de los riesgos y beneficios de los medicamentos promueve la seguridad del paciente.

## Funciones del sistema de farmacovigilancia

El sistema de farmacovigilancia debe cumplir con la **función primordial**, que es la **protección** de la población usuaria de medicamentos; para lograr lo anterior debe cumplir con los siguientes aspectos:

- **Detectar** eventos de baja ocurrencia.
- **Proveer información** que reafirme el nivel de seguridad de los medicamentos que actualmente son empleados en la población.
- **Detectar los aumentos o disminuciones** de los eventos adversos ya conocidos.
- **Identificar potenciales factores de riesgo** para las RAM.
- **Realizar una evaluación** sobre cada uno de los eventos adversos o RAM reportados al sistema de farmacovigilancia.
- **Determinar si existe o no relación causal** de los eventos o reacciones con el medicamento, biológico o vacuna bajo sospecha.

- Si es posible, **monitorizar** la seguridad entre lotes.
- **Retroalimentar** a la industria farmacéutica para la mejora de subproductos.
- **Retirar del mercado** productos no seguros para la población.
- En caso necesario, **promover las modificaciones** del registro en cuanto a lo indicado en las restricciones de uso de los mismos.

La seguridad de los medicamentos es una parte esencial de la seguridad de los pacientes. A nivel mundial, depende de la existencia de sistemas nacionales sólidos que permitan vigilar el desarrollo y la calidad de los medicamentos, informar sobre sus efectos perjudiciales y facilitar información precisa para su uso seguro.

Las RAM que se presenten con el uso de los medicamentos, sean éstas esperadas o inesperadas, deben ser notificadas; por ejemplo, reacciones:

- De fármacos de reciente introducción
- Que pongan en peligro la vida del paciente
- Que provoquen malformaciones congénitas
- Que originen ingreso hospitalario
- Que alarguen la estancia hospitalaria
- Que causen ausencia laboral
- Que provoquen efectos irreversibles
- Cualquiera que sea identificada por el profesional de la salud

No existen medicamentos exentos de riesgos y todos tienen efectos secundarios, algunos de los cuales pueden ser graves (cuadro 6-3). No es posible predecir con certeza absoluta los efectos de ningún tratamiento con medicamentos. Todo medicamento supone un compromiso entre los beneficios y los posibles perjuicios; estos últimos pueden reducirse al mínimo asegurando la buena calidad, la seguridad y eficacia del medicamento, su prescripción y uso racionales. Las RAM afectan a personas de todos los países del mundo y al menos 60% de ellas son evitables; sus causas pueden ser algunas de las siguientes:

**Cuadro 6-3.** Ejemplos de reacciones adversas a medicamentos (RAM)

| Medicamentos                                            | Reacciones                           |
|---------------------------------------------------------|--------------------------------------|
| Aminofenazona (antiinflamatorio)                        | Alteraciones de los glóbulos         |
| Clioquinol (antimicrobiano cutáneo)                     | Alteraciones visuales                |
| Estolato de eritromicina (antibacteriano)               | Hepatitis                            |
| Anticonceptivos orales                                  | Tromboembolias (coágulos sanguíneos) |
| Estatinas (control del colesterol)                      | Degeneración muscular                |
| Talidomida (control de las náuseas durante el embarazo) | Focomelia (malformación congénita)   |

- Diagnóstico erróneo.
- Prescripción del medicamento equivocado o de una dosis errónea del medicamento correcto.
- Trastornos médicos, genéticos o alérgicos subyacentes que pueden provocar una RAM.
- Automedicación con medicamentos que requieren prescripción.
- Incumplimiento del tratamiento prescrito.
- Reacciones con otros medicamentos (incluidos los medicamentos tradicionales) y determinados alimentos.
- Uso de medicamentos de calidad inferior a la norma, cuyos ingredientes y composición no se ajustan a las especificaciones científicas apropiadas, y que pueden resultar ineficaces y a menudo peligrosos.
- Uso de medicamentos falsificados sin ingredientes activos o con ingredientes equivocados, que pueden ser peligrosos.

## Clasificación de eventos adversos

Las sospechas de reacción adversa y las reacciones adversas de los medicamentos se clasifican, de acuerdo con la intensidad o gravedad de la manifestación clínica, en:

- **Leves.** Se presentan con signos y síntomas fácilmente tolerados; no necesitan tratamiento ni prolongan la hospitalización y pueden o no requerir de la suspensión del medicamento.
- **Moderadas.** Interfieren con las actividades sin amenazar directamente la vida del paciente. Requieren de tratamiento farmacológico y pueden o no requerir la suspensión del medicamento causante de la reacción adversa.
- **Graves (serias).** Cualquier manifestación que se presente con la administración de cualquier dosis de un medicamento y que conlleva alguno de los siguientes efectos:
  - Pone en peligro la vida o causa la muerte del paciente.
  - Hace necesario hospitalizar o prolongar la estancia hospitalaria.
  - Es causa de invalidez o de incapacidad permanente o significativa.
  - Es causa de alteraciones o malformaciones en el recién nacido.
- **Letal.** Contribuye de manera directa o indirecta a la muerte del paciente.

## Notificación individual de un caso

En farmacovigilancia, una notificación individual de un caso se puede definir como **una notificación relativa a un paciente que ha presentado un acontecimiento médico adverso (o alteración en pruebas de laboratorio) del que se sospecha está ocasionado por un medicamento.**

Una notificación individual de un caso debe contener (como mínimo para poder considerarse como tal) información de los siguientes aspectos:

- **El paciente.** Edad, sexo y breve historia clínica (cuando sea relevante). En algunos países es necesario especificar el origen étnico.
- **Acontecimientos adversos.** Descripción (naturaleza, localización, intensidad, características), resultados de investigaciones y pruebas, fecha de inicio, evolución y desenlace.
- **Fármacos sospechosos.** Nombre (marca comercial o nombre genérico del fármaco y fabricante), dosis, vía de administración, fechas de inicio y final del tratamiento.
- **Otros fármacos utilizados por el paciente** (incluyendo los de automedicación). Nombres, dosis, vías de administración, fechas de inicio y final.
- **Factores de riesgo** (p. ej., alteración de la función renal, exposición previa al fármaco sospechoso, alergias conocidas, uso de drogas sociales).
- **El nombre y la dirección del notificador** (debe considerarse confidencial y sólo utilizarse para verificar los datos, completarlos o hacer un seguimiento del caso).

## Medidas de seguridad

En todos los países la ley obliga a las empresas farmacéuticas, o fabricantes de medicamentos, a probar sus fármacos en voluntarios sanos y enfermos, antes de comercializarlos. Los ensayos clínicos muestran si el fármaco funciona y cuál es su eficacia para una determinada enfermedad, así como sus posibles efectos perjudiciales. Sin embargo, no proporcionan información sobre lo que ocurre en poblaciones más amplias con características distintas (edad, sexo, estado de salud, origen étnico, etc.) a las de los participantes en los ensayos clínicos.

La vigilancia de la seguridad de muchos medicamentos, y en particular de los productos complejos, no termina en la fase de fabricación. Debe continuar con una cuidadosa vigilancia de los pacientes y la recolección de más datos científicos. Este aspecto de la vigilancia de los medicamentos es lo que se denomina **vigilancia poscomercialización** o simplemente **farmacovigilancia**, y su eficacia a nivel nacional depende directamente de la participación activa de los profesionales de la salud.

Los profesionales de la salud (médicos, farmacéuticos, enfermeras, dentistas y otros) son quienes están en las mejores condiciones para informar cuando hay sospecha de reacciones adversas, como parte de la atención que prestan habitualmente a sus pacientes. Los profesionales de la salud deben informar sobre las RAM aun cuando tengan dudas sobre la relación precisa entre el medicamento en cuestión y la reacción.

## La farmacovigilancia en la práctica clínica

La vigilancia de la seguridad de los medicamentos es un elemento esencial para el uso eficaz de los mismos y para brindar una atención médica de alta calidad. La farmacovigilancia es una disciplina clínica por derecho propio que contribuye a la ética de la seguridad y sirve como un indicador de las normas de atención clínica practicadas en un

país. El control de la seguridad de los medicamentos de uso corriente debería ser parte integral de la práctica clínica. La medida en que el personal clínico está informado de los principios de la farmacovigilancia y ejerce su labor de acuerdo con ellos tiene gran incidencia en la calidad de la atención médica. Son elementos que redundan en una mejor atención al paciente la formación teórica y práctica del personal de salud sobre seguridad de los medicamentos; el intercambio de información entre centros nacionales de farmacovigilancia; la coordinación de esos intercambios, y la vinculación entre la experiencia clínica en este terreno, la investigación y la política sanitaria.

## Farmacovigilancia en México

En México, en 1989 se da inicio al Programa de Notificación Voluntaria de Sospechas de Reacciones Adversas a Medicamentos, con la participación de los laboratorios productores y la Secretaría de Salud a través de la Dirección General de Insumos para la Salud de la Secretaría de Salud; así se implementó el Sistema Nacional de Farmacovigilancia (actualmente COFEPRIS). En México, desde 2001 el COFEPRIS cuenta con el Centro Nacional de Farmacovigilancia (CNFV), en el cual los eventos adversos se reportan vía telefónica y mediante correo electrónico dirigido a la Dirección General de Epidemiología, de acuerdo a lo que está establecido en la NOM-220-SSA-2002, para la instalación y operación de la farmacovigilancia.

También desde el año 2001 se cuenta con el Centro Nacional de Farmacovigilancia (CNF) que forma parte de la Comisión Federal para la Protección contra Riesgos Sanitarios (COFEPRIS), el cual es el órgano rector en materia de farmacovigilancia en México, cuya finalidad es recibir informes sobre la detección de sospechas de reacciones adversas de los medicamentos, vacunas y dispositivos médicos, por parte de los profesionales de la salud, de los laboratorios productores, así como de los Centros Estatales de Farmacovigilancia ubicados en cada entidad federativa, con la finalidad de evaluar estas sospechas y retroalimentar la información, entre otros aspectos (figura 6-3.)

Además de seguir las recomendaciones internacionales para los sistemas de farmacovigilancia, cada país tiene un marco legal bien definido; en México se cuenta con la siguiente normatividad:

- Constitución Política de los Estados Unidos Mexicanos.
- Art. 58 V bis de la Ley General de Salud. *Diario Oficial* de la Federación, 7 de mayo de 1997.
- Art. 38 del Reglamento de insumos para la salud. *Diario Oficial* de la Federación, 4 de febrero de 1998.
- Art. 131 del Reglamento de insumos para la salud. *Diario Oficial* de la Federación, 4 de febrero de 1998.
- Reglamento de la COFEPRIS. *Diario Oficial* de la Federación, 13 de abril de 2004.
- NOM-220-SSA1-2002. Instalación y operación de la farmacovigilancia (15 de enero de 2005). *Diario Oficial* de la Federación, 15 de noviembre de 2004.



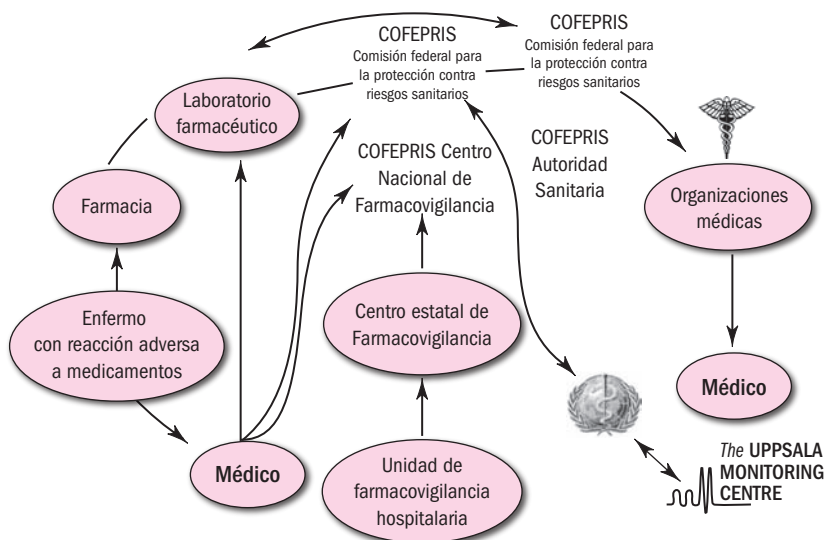


**Figura 6-3.** Flujo de información del proceso de farmacovigilancia en México.

- En México, la **notificación es obligatoria** en todo el territorio nacional para:
  - Instituciones que brindan servicios de salud.
  - Profesionales de la salud.
  - Titulares del registro sanitario.
  - Comercialización de medicamentos y productos herbolarios.
  - Unidades de investigación que realizan ensayos clínicos.

## Conclusiones

La farmacovigilancia es una actividad fundamental que se realiza para todos los medicamentos, biológicos, vacunas y dispositivos médicos, que se lleva a cabo una vez que éstos han pasado por las etapas de desarrollo, evaluación preclínica (en animales) y ensayos clínicos (fases I-III) y que ya han sido aprobados para su comercialización. La farmacovigilancia se realiza en la etapa de poscomercialización de los productos mencionados. Es importante destacar que la farmacovigilancia es una actividad en la que participan diversos actores, personal de salud, laboratorios productores, autoridades reguladoras (como la FDA) y centros de investigación donde se realizan ensayos clínicos, entre otros.

A nivel internacional, el programa de farmacovigilancia de la OMS implica proporcionar información sobre posibles aspectos de seguridad de medicamentos que aún no se hayan detectado en los países, con la finalidad de tomar medidas oportunas.

La farmacovigilancia es necesaria para la prevención de riesgos de los medicamentos en los seres humanos y para evitar los costos económicos asociados a las reacciones adversas a medicamentos.

## Referencias

1. Food and Drug Administration. A History of the FDA and Drug Regulation in the United States. <http://www.fda.gov/centennial/history/history.html> Consultado el 23 de febrero de 2012.
2. Secretaría de Salud. Comisión Federal para la Protección contra Riesgos Sanitarios. *Norma Oficial Mexicana NOM-059-SSA1-2006. Buenas prácticas de fabricación para establecimientos de la industria químico-farmacéutica dedicados a la fabricación de medicamentos*. Publicada en el 31 *Diario Oficial de la Federación* el 31 de julio de 1998.
3. OMS. Vigilancia de la seguridad de los medicamentos. *Guía para la instalación y puesta en funcionamiento de un centro de farmacovigilancia*. Publicada por: The Uppsala Monitoring Centre (UMC), WHO Collaborating Centre for International Drug Monitoring. 2001.
4. WHO. The Uppsala Monitoring Centre (UMC) <http://www.who-umc.org/> Consultado el 28 de febrero de 2012.
5. World Health Organization. The Importance of pharmacovigilance. (Safety monitoring of medicinal products.) 2002.
6. Red Panamericana de Armonización de la Reglamentación Farmacéutica. *Buenas prácticas de farmacovigilancia para las Américas*. 2008.
7. OMS. *Medicamentos: seguridad y reacciones adversas*. Nota descriptiva N° 293. Octubre de 2008. <http://www.who.int/mediacentre/factsheets/fs293/es/index.html>
8. OMS. *Perspectivas políticas de la OMS sobre medicamentos. La farmacovigilancia: garantía de seguridad en el uso de los medicamentos*. Ginebra. Octubre de 2004.
9. *Norma Oficial Mexicana NOM-220-SSA1-2004. Instalación y operación de la farmacovigilancia*. *Diario Oficial de la Federación*. Fecha de publicación: 15 de noviembre de 2004.
10. Talbot JCC, Nilsson BS. *Pharmacovigilance in the pharmaceutical industry*. *BJl of Pharmacology* 45:427-431. 1998.
11. Secretaría de Salud. *Farmacopea de los Estados Unidos Mexicanos. Características que debe cumplir un medicamento para obtener el Registro Sanitario*, 8a. ed. Vol. 1;20-34. 2004.
12. Secretaría de Salud. Comisión Nacional para la Protección contra Riesgos Sanitarios. <http://www.cofepris.gob.mx/AZ/Paginas/Farmacovigilancia>
13. World Health Organization. Uppsala Monitoring Centre. Glossary of terms used in Pharmacovigilance. August 2011.



# Capítulo 7

## Bioética y Epidemiología Clínica

*Horacio García Romero  
Luis Limón Limón*

### Introducción

La investigación en salud es uno de los capítulos más importantes de la Bioética, la cual puede definirse como el estudio sistemático de las normas y acciones morales de todos los individuos relacionados con la ciencia de la vida y la atención a la salud.

Las investigaciones que se realizaron durante la Segunda Guerra Mundial en seres humanos, muchas de las cuales pueden considerarse como contrarias a la razón y a la moralidad, obligaron a los países a establecer normas que fijaran las bases de toda investigación clínica. Los juicios de Núremberg (1947) concluyeron en declaraciones muy estrictas. En 1964, la Asamblea Médica Mundial estableció en la Declaración de Helsinki los principios básicos que fueron aceptados por la mayoría de las naciones en el mundo e incorporados en leyes y reglamentos de cada uno de los países.

En México, tanto la *Ley General de Salud* como el *Reglamento en materia de investigación para la salud* incorporaron muchos de los conceptos de la Declaración de Helsinki, lo que les dio consistencia y una fuerza poco común.

La *Ley General de Salud* señala que toda institución en la que se realicen investigaciones en seres humanos debe contar por lo menos con dos comités que autoricen la realización de cualquier investigación:

1. Un Comité científico, que compruebe la validez del estudio, que dé el visto bueno al protocolo y que verifique el seguimiento de los pacientes o individuos involucrados.
2. Un Comité de ética, que ratifique que la investigación no salga de las normas aceptadas. Además, si en la situación se estudian elementos vivos patógenos (bacterias, virus) o se llevan a cabo investigaciones genéticas, se debe establecer:
3. Un Comité de bioseguridad. La Secretaría de Salud ha promulgado un *Reglamento en materia de investigación para la salud* que incluye un Título Segundo: “De los preceptos éticos de la investigación en seres humanos”, que contiene cinco capítulos.

El *Reglamento de investigación* de la *Ley General de Salud de México* señala puntos muy relevantes en referencia a principios éticos que deben regular las investigaciones e incluso hace mención de muchas bases científicas y técnicas que les dan valor a las investigaciones en seres humanos.

En el artículo 14 se anota que la investigación contará con el dictamen favorable de los Comités de Investigación, Ética y, en su caso, Bioseguridad.

En el artículo 15 se señala que cuando el diseño experimental de una investigación incluya varios grupos, se usarán métodos aleatorios de selección para obtener una asignación imparcial de los participantes en cada grupo.

En el artículo 17 las investigaciones se clasifican según la probabilidad de que el sujeto de investigación sufra algún daño como consecuencia del estudio.

1. **Investigación sin riesgo.** Son estudios que emplean técnicas y métodos de investigación documental retrospectivos y aquellos en los que no se realiza ninguna intervención o modificación intencionada en las variables fisiológicas, psicológicas y sociales de los individuos que participan en el estudio, entre los que se consideran: cuestionarios, entrevistas, revisión de expedientes clínicos y otros, en los que no se identifiquen ni se traten aspectos sensitivos de su conducta.
2. **Investigación con riesgo mínimo.** Estudios prospectivos que emplean el registro de datos a través de procedimientos comunes en exámenes físicos o psicológicos de diagnóstico o tratamientos rutinarios, como son: pesar al sujeto, pruebas de agudeza auditiva, electrocardiograma, termografía, colección de líquido amniótico al romperse las membranas, extracción de sangre por punción venosa en adultos en buen estado de salud, pruebas psicológicas a individuos o grupos en los que se manipulará la conducta del sujeto, entre otras.
3. **Investigación con riesgo mayor que el mínimo.** Son aquellas en que las probabilidades de afectar al sujeto son significativas; entre ellas se consideran: estudios radiológicos y con microondas, ensayos con los medicamentos con los nuevos dispositivos, estudios que incluyen procedimientos quirúrgicos, extracción de sangre mayor al 2% del volumen circulante en neonatos, amniocentesis y otras técnicas invasoras o procedimientos mayores, los que empleen métodos aleatorios de asignación a esquemas terapéuticos y los que tengan control con placebos, entre otros.

Esta clasificación reviste una particular importancia al considerar el consentimiento informado que deben otorgar los individuos sujetos a la investigación, como se señala en el artículo 23 del propio *Reglamento*.

En caso de investigaciones con riesgo mínimo, la Comisión de Ética, por razones justificadas, podrá autorizar que el consentimiento informado se obtenga sin formularse por escrito y, tratándose de investigaciones sin riesgo, podrá dispensar al investigador la obtención del consentimiento informado.

Cuando el consentimiento informado se formule por escrito, debe llenar los requisitos siguientes (artículo 22):

- Será elaborado por el investigador principal.

- Será revisado y, en su caso, aprobado por la Comisión de Ética de la institución de atención a la salud.
- Indicará los nombres y direcciones de dos testigos y la relación que éstos tengan con el sujeto de investigación.
- Deberá ser firmado por dos testigos y por el sujeto de investigación o su representante legal, en su caso. Si el sujeto de investigación no sabe firmar, es indispensable que imprima su huella digital y que a su nombre firme otra persona que él designe. Estos documentos se extenderán por duplicado y uno de los ejemplares debe quedar en poder del sujeto de investigación o de su representante legal.

El artículo 21 expone ampliamente lo que debe conocer el sujeto de investigación o, en su caso, su representante legal, mediante explicaciones claras y completas de los siguientes puntos:

- La justificación y los objetivos de la investigación.
- Los procedimientos que vayan a usarse y su propósito, incluyendo la identificación de los procedimientos que son experimentales.
- Las molestias o riesgos esperados.
- Los beneficios que puedan obtenerse.
- Los procedimientos alternativos que pudieran ser ventajosos para el sujeto.
- La garantía de recibir respuesta a cualquier pregunta y aclaración a cualquier duda acerca de los procedimientos, riesgos, beneficios y otros asuntos relacionados con la investigación y el tratamiento del sujeto.
- La libertad de retirar su consentimiento en cualquier momento y de dejar de participar en el estudio, sin que por ello se creen perjuicios para continuar su cuidado y tratamiento.
- La seguridad de que no se identificará al sujeto y de que se mantendrá la confidencialidad de la información relacionada con su privacidad.
- El compromiso de proporcionar al individuo la información actualizada obtenida durante el estudio, aunque ésta pudiera afectar la voluntad del sujeto para continuar participando.
- La disponibilidad de tratamiento médico y la indemnización a que legalmente tendría derecho por parte de la institución de atención a la salud, en el caso de daños que la ameriten, directamente causados por la investigación.
- La certeza de que si existen gastos adicionales, éstos serán absorbidos por el presupuesto de la investigación.

El *Reglamento* incluye dos capítulos (III y IV) que aluden a la investigación en menores, en individuos con incapacidad mental y en mujeres embarazadas. El artículo 36 indica que para la realización de investigaciones en menores o incapaces se deberá, en todo caso, obtener el consentimiento informado por escrito de quienes ejerzan la patria potestad o la representación legal del menor o incapaz de que se trate.

Cuando dos personas ejerzan la patria potestad de un menor, sólo será admisible el consentimiento de una de ellas si existe imposibilidad fehaciente o manifiesta de la otra para proporcionarlo o en caso de riesgo inminente para la salud o la vida del menor incapaz.

Cuando la capacidad mental y estado psicológico del menor o incapaz lo permitan, deberá obtenerse, además, su aceptación para ser sujeto de investigación; después de explicar lo que se pretende hacer, el Comité de Ética podrá dispensar el cumplimiento de estos requisitos por razones justificadas (artículo 37).

En los artículos 43 a 47 se señala que para realizar investigaciones en mujeres embarazadas, durante el trabajo de parto, puerperio y lactancia; en nacimientos vivos o muertos; de utilización de embriones, óbitos o fetos, y para la fertilización asistida, se requiere obtener la carta de consentimiento informado de la mujer y de su cónyuge o concubinario, previa información de los riesgos posibles para el embrión, feto o recién nacido en su caso.

El consentimiento del cónyuge o concubinario sólo podrá dispensarse en caso de incapacidad o imposibilidad fehaciente o manifiesta para proporcionarlo; porque el concubinario no se haga cargo de la mujer, o bien cuando exista riesgo inminente para la salud o la vida de la mujer, embrión, feto o recién nacido.

Las investigaciones sin beneficio terapéutico sobre el embarazo no deberán representar un riesgo mayor al mínimo para la mujer, el embrión o el feto.

Las investigaciones en mujeres embarazadas que impliquen una intervención o procedimiento experimental no relacionado con el embarazo, pero con beneficio terapéutico para la mujer, como sería en casos de diabetes, hipertensión y neoplasias, entre otros, no deberán exponer al embrión o al feto a un riesgo mayor al mínimo, excepto cuando el empleo de la intervención o procedimiento se justifique para salvar la vida de la mujer.

Las investigaciones en mujeres embarazadas, con beneficio terapéutico relacionado con el embarazo, se permitirán cuando:

- I. Tengan por objeto mejorar la salud de la embarazada con un riesgo mínimo para el embrión o feto, o
- II. Estén encaminadas a incrementar la viabilidad del feto, con un riesgo mínimo para la embarazada.

En el capítulo V del *Reglamento* se dan indicaciones sobre la investigación en grupos subordinados como estudiantes, empleados en un hospital, miembros de las fuerzas armadas, internos en reclusorios o centros de readaptación social, en los que el consentimiento informado pueda ser influido por alguna autoridad. En todos los estudios con grupos subordinados, el Comité de Ética debe tener especial cuidado en que el rechazo a intervenir o el retiro del consentimiento no afecte la situación laboral, escolar, militar o judicial de los individuos.

## Requisitos para que una investigación clínica sea ética

Los requisitos más importantes para que una investigación clínica sea ética han sido resumidos por Ezequiel Emanuel, quien hace énfasis en que el consentimiento informado, aunque necesario, no es el más trascendental de esos requisitos.

Las siete condiciones en orden de importancia, según este autor, para que una investigación clínica sea ética son las siguientes:

1. Que sea valiosa; que tenga importancia social, científica o clínica. Que aporte información que, en la práctica, sea beneficiosa para los enfermos o para la sociedad. De esta manera se asegura que los sujetos de investigación no sean expuestos a riesgos sin la posibilidad de algún beneficio personal o social.
2. Que tenga validez científica. El protocolo de la investigación debe seguir las normas científicas que exige un buen diseño. Un estudio en sujetos humanos indebidamente diseñado, que no tiene posibilidad de producir hechos científicos (observaciones reproducibles), no es ético. La metodología del estudio debe tener sólidas bases científicas y ser prácticamente realizable. Sólo así se justifican los riesgos de cualquier investigación clínica.
3. Que los individuos del grupo o de los grupos en estudio hayan sido seleccionados de manera equitativa. No es correcto seleccionar individuos menos capaces de hacer valer sus derechos (niños con retraso mental, personas en extrema pobreza o con gran ignorancia) si no tienen una finalidad directa en la investigación.

Una selección equitativa de sujetos requiere también que a todos los grupos se les ofrezca la oportunidad de participar en el estudio, salvo si existen razones científicas que lo impidan.

Además, la selección de sujetos puede considerarse equitativa sólo cuando aquellos que se reclutan están en condiciones de beneficiarse si la investigación proporciona un resultado positivo, como puede ser un nuevo tratamiento.

Por último, la selección equitativa de los sujetos debe estar relacionada con el riesgo-beneficio que se espera en cada caso. Pueden excluirse aquellos individuos en los que el riesgo de agravarse es grande, o en quienes el beneficio esperado será mucho menor.

4. Que la investigación que se realiza tenga, en general, una probabilidad de obtener beneficios mucho mayores que los riesgos que puedan resultar del estudio. Esto puede no ser fácil de definir. En ocasiones un estudio piloto previo con un número pequeño de sujetos ayuda a tomar una decisión sobre la factibilidad ética de una investigación.
5. Que la evaluación de los datos en los individuos estudiados se realice por personas expertas que no estén directamente involucradas en la investigación (evaluación independiente). Los investigadores, por diversas razones de prestigio, de deseos de contribuir a la ciencia y otros, tienden, sin mala fe, a encontrar en sus estudios información cuyo valor no tiene suficiente fundamento. La evaluación independiente evita que esto ocurra.
6. Que se tenga el consentimiento informado de los sujetos de la investigación en todas aquellas que tengan un riesgo mayor que el mínimo. El *Reglamento de la Ley General de Salud* señala con detalle las variantes de esta norma, que ya fueron expuestas.
7. Que antes, durante y después del estudio se tenga respeto o interés genuino en el bienestar de los sujetos involucrados; el consentimiento informado no termina con la firma del documento por el sujeto en el que se investiga. El seguimiento continuo

del proceso, el permitir que cualquier individuo pueda retirarse del estudio si así lo desea, el informarle sobre el aumento de los riesgos durante la investigación constituyen también obligaciones éticas de los investigadores.

Así, se reconoce que la dignidad, los intereses y el bienestar de los pacientes deben prevalecer sobre cualquier otra meta fijada en la investigación.

La investigación clínica en seres humanos es un tema de sumo interés para la Bioética. En efecto, el avance tecnológico en los métodos de tratamiento y de diagnóstico repercute de manera incuestionable en la salud de un individuo y de la población en general. El personal de salud debe procurar involucrarse en las investigaciones que se realicen en su institución y para ello necesita conocer cómo llevarlas a cabo y cuáles son las normas científicas y éticas en las que se sustenta.

Realizar una investigación clínica que no cumpla con los preceptos científicos no sólo produce un daño directo en cuanto a gasto inútil del tiempo de los investigadores, sino desperdicio de recursos y riesgo para los pacientes. Puede, además, llevar a conclusiones falsas, las que al divulgarse incrementan el daño; de ahí que se consideran faltas de ética.

La Bioética no sólo contribuye a fijar las normas para la investigación en salud, pues es una ciencia que, a su vez, abarca temas que deben investigarse; los siguientes son algunos ejemplos:

- a) Calidad de la atención a la salud.
- b) Avances tecnológicos y su utilización adecuada.
- c) Asignación de recursos.
- d) Instituciones de seguros especializadas en salud.
- e) Leyes, reglamentos, normas y programas sobre atención a la salud. Vigencia y cumplimiento.
- f) Actividades del personal de salud. Cumplimiento de las normas bioéticas.
- g) Manejo de conflictos específicos al principio y al final de la vida en las instituciones.
- h) Vigencia y cumplimiento de leyes y reglamentos en relación con el entorno ambiental.
- i) Conocimientos y opiniones del personal de salud y de la comunidad sobre temas bioéticos.

La relación de los médicos con los representantes de la industria farmacéutica ha llegado a ser muy estrecha, y el profesional se siente obligado a retribuir a esa industria los regalos, las becas, los obsequios económicos durante una investigación y, como consecuencia, los datos que se reportan sobre productos nuevos pueden desviarse de la realidad. Lo anterior constituye un problema que denigra a la profesión médica y que obliga cada vez más a recapacitar sobre la integridad, la honestidad y el cuidado con el que debe tratar a sus pacientes, así como a la necesidad de que el investigador realice su trabajo siempre con apego a la verdad, así sea en contra de sus propios intereses. La bioética adquiere gran relevancia en esta fase del comportamiento del médico.

En la Declaración de Helsinki adoptada por la Asamblea Médica Mundial en 1968 se indican los Principios éticos para las investigaciones médicas en seres humanos. Este documento ha sido la base, en la mayoría de los países en el mundo, para el estableci-



miento de las normas que deben seguir los investigadores clínicos. La Declaración se revisó en Tokio, Japón (1975), Venecia, Italia (1983), Somerset West, Sudáfrica (1996) y Edimburgo, Escocia (2000). En 2002 se agregó en Washington una “Nota de clarificación del Párrafo 29”, que trata sobre el uso de placebos. Esta “clarificación” ha sido gravemente criticada porque permite el empleo de sustancias inertes (placebos) en el grupo testigo (control) de una investigación, aun cuando se conocen ya medicamentos activos que alivian o que curan una enfermedad grave. Esto se ha hecho en África para probar sustancias en el tratamiento de gestantes con SIDA, con intención de evitar el paso del virus al producto del embarazo. Las mujeres que reciben placebo no pueden evitar la transmisión de esta enfermedad a sus hijos por nacer.

En la Asamblea Médica Mundial y otras reuniones internacionales, el hecho de tratarse de opiniones de expertos no impide que éstos se inclinen por intereses privados o de grupos poderosos, ajenos a los preceptos sociales de la Bioética. Aquellos que genuinamente se interesan por la salud y por el bien de todos deben estar atentos a que los representantes de los países desarrollados no apliquen una bioética de “doble estándar”, una para los ricos y otra para los pobres.

---

## Bibliografía

- Cañedo L, García Romero H, Méndez R. *Principios de investigación médica*, IMAN, México, 1980.
- Código Internacional de Ética*, Asamblea Médica Mundial, OMS, Ginebra, Suiza, 1949.
- Emanuel E J. The relevance of empirical research in Bioethics, en OMS OPS, Publication Series, 2002.
- García Romero H, García-Procel E. La ética médica profesional y las prácticas de las industrias farmacéuticas. *Gaceta Médica de México* 142:439-440. 2006.
- García-Romero H, Limón LL. *Bioética general*. Trillas, México. 2009.

# Capítulo 8

## Revisiones sistemáticas y metaanálisis en Medicina

*Patricia Clark  
Rodolfo Rivas Ruiz*



### Introducción

La revisión sistemática constituye una estrategia de revisión estructurada de la literatura científica que surge de una pregunta clínica e intenta resolver las dudas que existan sobre esta pregunta específica. Esta estrategia se encuentra clasificada como un estudio de investigación secundaria;<sup>1</sup> esto quiere decir que utiliza a los estudios clínicos originales para su análisis; por ejemplo, en una paciente con dermatomiositis activa, ¿cuánto tiempo se debe sostener el tratamiento con dosis altas de esteroides?, ¿cuál es el pronóstico con este tratamiento? Las decisiones clínicas deben estar basadas en la mejor evidencia proporcionada por estudios reportados en la literatura científica, realizados en grupos de pacientes similares al caso en cuestión, y de los cuales es posible tener la certeza de que dichos resultados puedan ser aplicables.

En la literatura científica es posible encontrar un cúmulo de estudios en todas las especialidades y enfermedades, pero ¿cuán contundentes son?; ¿qué tan bien fueron realizados?; ¿se parece el grupo de pacientes del estudio a los pacientes en cuestión?; ¿es posible aplicar esos resultados en tal o cual práctica clínica diaria, o en el Sistema de Salud de un país determinado?

El objetivo de las revisiones sistemáticas es proporcionar a los clínicos una respuesta objetiva para la toma de decisiones basada en el resumen de todas las evidencias disponibles. Para lograr este invaluable producto científico se valora la calidad de los estudios, la manera en que fueron realizados y los estándares metodológicos. Con estos parámetros es posible calificar de forma objetiva los resultados para ver la pertinencia de los mismos y la aplicabilidad en los pacientes determinados; cuando esta revisión, además de tener el resumen cualitativo de la calidad de los estudios originales, cuenta también con un resumen numérico derivado de la combinación matemática de los datos, recibe el nombre de **metaanálisis**.

## Caso clínico

---

Una mujer de 34 años de edad, con el diagnóstico de fibromialgia con dos años de evolución, acude a consulta médica y refiere que a pesar del tratamiento con dosis adecuadas de pregabalina desde hace siete meses, persisten algunos síntomas propios de la enfermedad que afectan su calidad de vida de acuerdo a la escala de gravedad de los síntomas. La paciente señala que el dolor generalizado ha disminuido, pero la fatiga y los síntomas de depresión se presentan con mucha frecuencia. En la exploración física se encuentran siete puntos en gatillo positivos además de lo referido en su historia clínica. Otras pacientes le han dicho que el ejercicio le puede ayudar y quiere saber si esta información es cierta, y de ser así qué tipo de ejercicio es recomendable.

Usted ha leído que el ejercicio tiene algún efecto positivo en estos casos, pero no está seguro de cuál es el tipo de ejercicio y cuál es la magnitud del efecto de esta intervención en la fibromialgia, por lo que se decide a hacer una revisión de la literatura científica para hacer una recomendación basada en la mejor evidencia.

---

## Revisiones narrativas vs. revisiones sistemáticas

Tradicionalmente, los estudios de revisión o integración de la literatura médica se han realizado de manera simple, a través de narraciones donde un experto en el tema revisa los estudios originales, decide cuáles son relevantes y resume de manera básica las conclusiones y los resultados y, en menor grado, los aspectos metodológicos. Por último, algunas de estas revisiones proponen futuras líneas de investigación sobre el tema. Este tipo de revisiones, predominantemente narrativas, se caracterizan por algunas propiedades que en la actualidad representan más problemas que virtudes para la toma de decisiones respecto a dilemas en la clínica, como son:

1. Por lo general estas revisiones plantean preguntas amplias (como “¿cuál es el manejo del paciente con enfermedad de Alzheimer?”).
2. Los autores revisan diferentes aspectos del manejo y la enfermedad teniendo como parámetros sus propias preferencias, mismas que pueden incluir por igual la fisiología, los criterios diagnósticos, exámenes diagnósticos, diferentes alternativas de tratamiento y hasta el pronóstico. La falta de estructura planteada *a priori* en la revisión de la literatura da como resultado que las recomendaciones en términos generales sean basadas en su experiencia y en los artículos que les parecen relevantes, más que en la evidencia de toda la literatura.
3. Estas revisiones, por su carácter subjetivo, no tienen reglas formales para realizarse, no utilizan una forma sistemática de localización y recolección de los estudios individuales, y existen sesgos que pueden ser introducidos por el revisor (de manera consciente o inconsciente) quien, a través de su juicio, elige los estudios para fundamentar sus conclusiones. Como consecuencia, no hay estándares explícitos o criterios bien definidos para evaluar estos estudios de investigación original y, con frecuencia, en diferentes revisiones sobre el mismo tema no existe consenso entre los distintos autores.

4. Por la inconsistencia de los métodos, falta de criterios y metodología precisa, se ignoran aspectos importantes como el tamaño de la muestra, la magnitud del efecto y el diseño de la investigación de los estudios originales. Esta forma de revisión es ineficaz para extraer información útil de los estudios, en especial cuando el número de artículos sobre el mismo tema es amplio.
5. No se sintetizan los datos cuantitativos, sólo se usa la significancia estadística o intervalos de confianza de los estudios individuales, de manera que la revisión es básicamente narrativa y no presenta ningún resumen numérico.

Debido a lo anterior, estas revisiones narrativas se encuentran sujetas a múltiples sesgos, y difícilmente permiten tomar decisiones clínicas basadas en ellas. Su utilidad reside en tener un resumen amplio y actualizado de algún tema, como puede ser algún capítulo de libro. Se deben utilizar como primer contacto de un tema. Por desgracia, este tipo de revisiones son ampliamente socorridas por los clínicos en búsqueda de respuestas para establecer el tratamiento en sus pacientes.

Las revisiones sistemáticas, a diferencia de las narrativas, son revisiones rigurosas con preguntas clínicas específicas y enfocadas (como “¿cuál es la efectividad del ácido zolendrónico en la disminución de fracturas por fragilidad en pacientes con osteoporosis?”). Son “sistemáticas” porque hacen un resumen de la investigación original del tema basándose en una metodología que obedece a un plan o **protocolo** escrito antes de iniciar la revisión; es decir, hacen explícitos cada uno de los pasos para su realización, de forma tal que los lectores puedan cuantificar la veracidad y reproducibilidad de los métodos empleados para hacer la revisión, y de esa forma, evaluar la validez de sus resultados y sus conclusiones. En la figura 8-1 se presenta un esquema de las características de ambas revisiones.

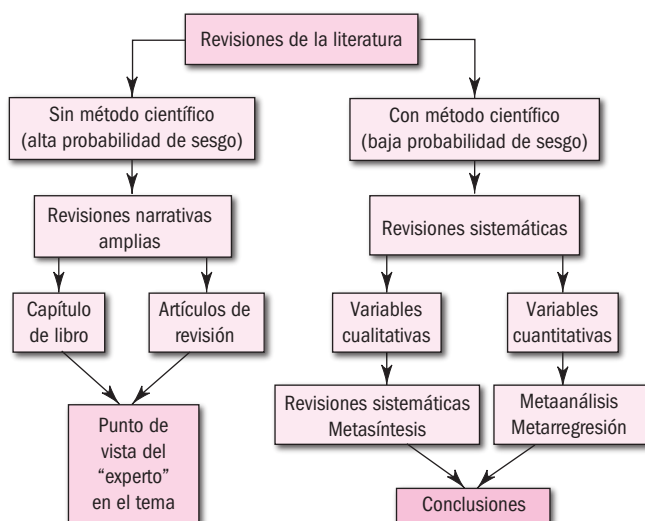


**Figura 8-1.** Tipos de revisiones en Medicina.

## ¿Cuándo es necesario tener una revisión sistemática de algún tema?

Las revisiones sistemáticas y los metaanálisis son muy útiles cuando se requiere responder a una pregunta específica, como “¿La actividad física reduce la mortalidad en pacientes con enfermedad cardiovascular previa? (infarto al miocardio, posoperados de revascularización de arterias coronarias, angioplastia coronaria transluminal, angina de pecho o enfermedad coronaria definida por angiografía)” o “¿Cuáles son los riesgos y beneficios de las dosis bajas de ácido acetilsalicílico para la prevención primaria de la enfermedad vascular (infarto al miocardio, accidente vascular cerebral o muerte por enfermedad vascular)?”

Además de tener una pregunta bien definida, es importante tener varios artículos originales del tema donde exista alguna controversia. No tiene caso realizar una revisión sistemática y metaanálisis si los estudios publicados son suficientes y todos consistentes en sus hallazgos. Este tipo de revisiones son útiles cuando los estudios originales no pueden contestar en forma adecuada su pregunta por alguna de las siguientes razones: estudios negativos por probable muestra pequeña (falta de poder), controversia en los resultados o conclusiones de estudios que abordan la misma pregunta, o en algunas ocasiones estudios que dejan sus preguntas abiertas.

En su inicio, el metaanálisis en Medicina se enfocó sólo en resumir los hallazgos de los ensayos clínicos controlados para la búsqueda de conclusiones en cuanto a efectividad terapéutica o efectos indeseables. Con el tiempo se han desarrollado nuevas técnicas metaanalíticas para realizar revisiones sistemáticas y metaanálisis en estudios observacionales con otros diseños como cohortes o casos y controles, que permiten combinar los datos de los factores de riesgo, como en el caso del cáncer de mama, o bien en los estudios de pruebas diagnósticas combinando la sensibilidad y especificidad.

## Importancia de las revisiones sistemáticas y metaanálisis en la Medicina moderna

Las revisiones sistemáticas y metaanálisis son estudios de integración, que han ganado popularidad en las últimas décadas, ya que la complejidad, el extenso volumen de la literatura científica y los resultados, a veces contradictorios, sobre los mismos tópicos, hacen difícil realizar la toma de decisiones clínicas. Como se aprecia en la figura 8-2, en el lustro de 1980 a 1984 se registraron 54 revisiones sistemáticas y metaanálisis en la base de datos de PubMed; para el lustro de 2005 a 2009 se reportaron más de 20 000 revisiones sistemáticas y metaanálisis. Sin embargo, este crecimiento acelerado no asegura que la calidad de los mismos sea óptima. Con el propósito de mantener una buena calidad en estos estudios, la colaboración Cochrane ha establecido criterios rígidos para la realización de revisiones sistemáticas, lo cual se aborda más adelante.

Este incremento en las publicaciones va de la mano con la difusión de estos estudios, así como su uso por parte de los médicos, investigadores y los tomadores de decisiones.

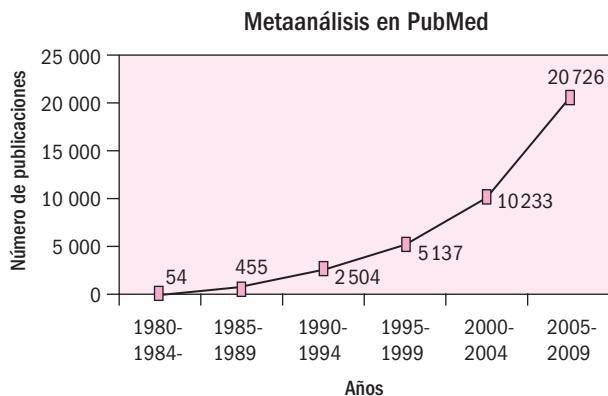


**Figura 8-2.** Incremento exponencial en el número de metaanálisis publicados y citados en PubMed de 1980 al 2009.

Un ejemplo de la importancia que tienen las revisiones sistemáticas radica en que no sólo son de utilidad para la toma de decisiones frente a un paciente, sino que forman parte del arsenal para la toma de decisiones en los sistemas de salud. Recientemente se ha propuesto el término de **“atención de la salud basada en la evidencia”** debido a la necesidad de implementar tratamientos o medidas de salud que sean eficientes y de mejor calidad.

## Utilidad de las revisiones sistemáticas

La mejor síntesis de la literatura médica es esencial para tomar decisiones en todas las áreas relacionadas con la salud: *a)* en la clínica, es indispensable para establecer el tratamiento para los pacientes en forma individual, en la búsqueda de nuevas opciones terapéuticas o dosis óptimas de tratamiento; *b)* en la realización de las guías de tratamiento de enfermedades comunes; *c)* en la investigación para generar nuevas hipótesis que se deriven de información válida y confiable; *d)* en el campo de la administración y planeación de la salud la mejor información médica es necesaria para la realización de programas costo-eficacia en las comunidades, así como para establecer políticas en la relación con la atención a la salud. La síntesis adecuada de la información es vital para la planeación de mejores estudios en la búsqueda de factores etiológicos de enfermedad o para el control de la diseminación de padecimientos.

En la figura 8-3 se observa el diseño de la secuencia lógica de este tipo de estudios que a continuación serán explicados.

## Definición del problema y objetivos

El primer paso, como en cualquier estudio de investigación, consiste en definir el problema que se desea estudiar al plantear la pregunta de investigación; por ejemplo, “¿El

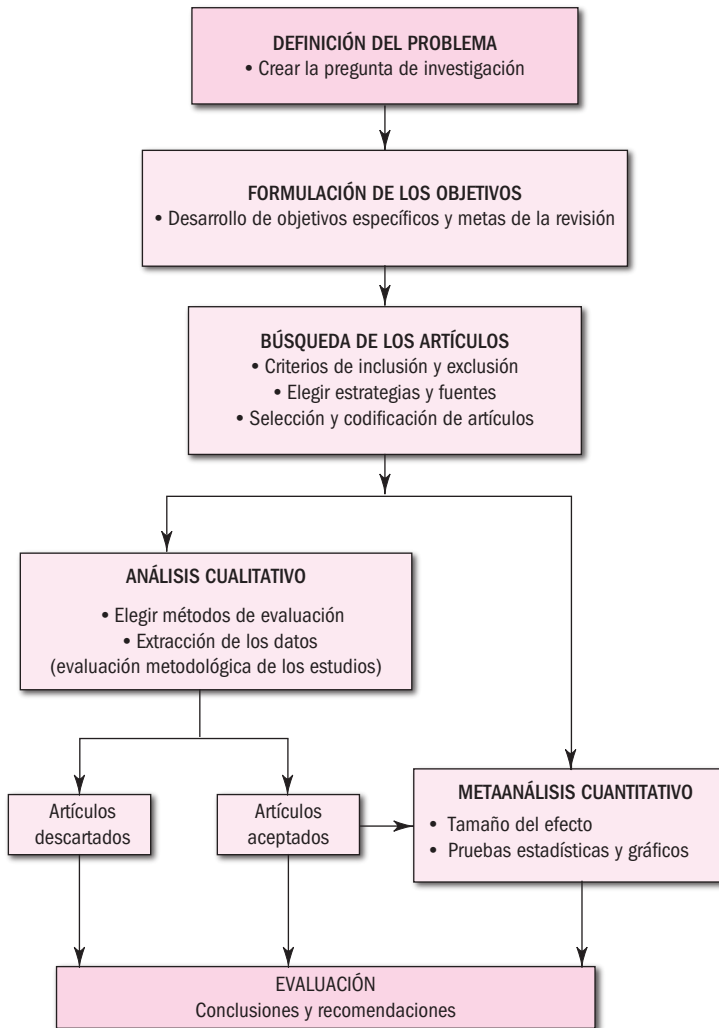


**Figura 8-3.** Flujograma del metaanálisis en Medicina.

tabaquismo activo disminuye la densidad ósea mineral?” o “¿La actividad física reduce las concentraciones de hemoglobina glucosilada en aquellos pacientes con diabetes mellitus tipo 2?” Como es evidente en estos ejemplos, las preguntas deben de ser claras y enfocadas. De la misma forma deben formularse el o los objetivos de la revisión.

## Búsqueda de los artículos y bases de datos electrónicas

La estrategia de búsqueda es uno de los pasos más importantes en las revisiones sistemáticas. Esta estrategia debe ser lo suficientemente clara y extensa para garantizar que se van

a encontrar todos los estudios para resolver la pregunta de investigación y no sólo una muestra sesgada de los mismos.

El primer paso es el definir los criterios de inclusión y exclusión, tales como el tema, el tipo de diseño, idioma, tipo de población (p. ej., adultos, niños, mujeres posmenopáusicas) y palabras clave que se utilizarán.

El segundo paso es elegir la estrategia y fuentes a revisar. La búsqueda debe iniciar en las bases electrónicas. Medline (PubMed) es una base de datos que reúne varias bases de datos de literatura científica. Tiene la ventaja de contener una amplia gama de artículos y ser gratuita. Esta base tiene un índice de palabras clave llamada palabras MeSH (del inglés, *Medical Sub Headlines*) que son de gran utilidad para la búsqueda. El principal competidor de Medline es EMBASE (*Excerpta Medica Data BASE*); es una base de datos bibliográfica producida por la empresa Elsevier. Es la versión electrónica del conocido índice Excerpta Medica, la cual es una base de datos primordialmente europea. Existen otras bases de datos especializadas de disciplinas afines como son: HealthSTAR, Psyc INFO y la biblioteca Cochrane (en sus versiones en inglés y español). Para América Latina es importante consultar las bases de datos que indexan a revistas que no son publicadas en las bases ya mencionadas; estas bases son auspiciadas por BIREME, el cual es un Centro Especializado de la Organización Panamericana de la Salud (OPS), establecido en Brasil desde 1967 y contiene bases como LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud), que alcanzó el 22 de noviembre de 2009 la marca de medio millón de documentos indizados. Es el índice bibliográfico de referencia de la producción científica y técnica en salud de la región de América Latina y el Caribe.

Existen otros buscadores, como Ovid y Science Direct, que requieren inscripción y son costosos; no obstante, tienen la ventaja de que se pueden obtener los artículos en extenso, pero tienen el inconveniente de que tienen un número limitado de revistas que depende del usuario que las contrata.

De manera adicional, se debe realizar una búsqueda manual de libros de texto recientes, revisiones tradicionales, buscar a los expertos en el área y revisar los registros de los ensayos clínicos, como los que se encuentran en CENTRAL de la colaboración Cochrane y *Clinical trials* ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)) para asegurar la inclusión de toda la información relevante.

Finalmente ha de intentarse una búsqueda de la literatura gris que comprende estudios que no han sido publicados en revistas por pares, documentos gubernamentales, trabajos presentados en congresos, tesis y disertaciones, que hubieran sido mostrados en otros foros científicos (reuniones y congresos) o sitios de Internet. Estos últimos se pueden hallar en buscadores como memorias de congresos o metabuscadores en Google o Yahoo.

## Selección de los artículos

Con la lista generada por la búsqueda electrónica, los artículos seleccionados serán revisados por lo menos por dos investigadores independientes (llamados pares) de acuerdo



a los criterios de inclusión y exclusión definidos previamente. Una vez terminada esta selección se realiza una prueba de concordancia entre los revisores mediante la prueba de correlación de kappa (figura 8-4) para validar la consistencia de la selección de artículos. En los estudios en donde no existe acuerdo, éste debe resolverse por consenso entre los revisores o por un tercer revisor. En la figura 8-4 se observa la secuencia que se llevó a cabo en una búsqueda sistemática que incluyó todos estos pasos.

Es importante describir el número de artículos que se incluyeron en la evaluación inicial, los que se excluyeron, así como las razones de la exclusión; por ejemplo, si en los criterios de inclusión la revisión será restringida a artículos en inglés y español, cualquier artículo en otro idioma quedará excluido de la revisión.

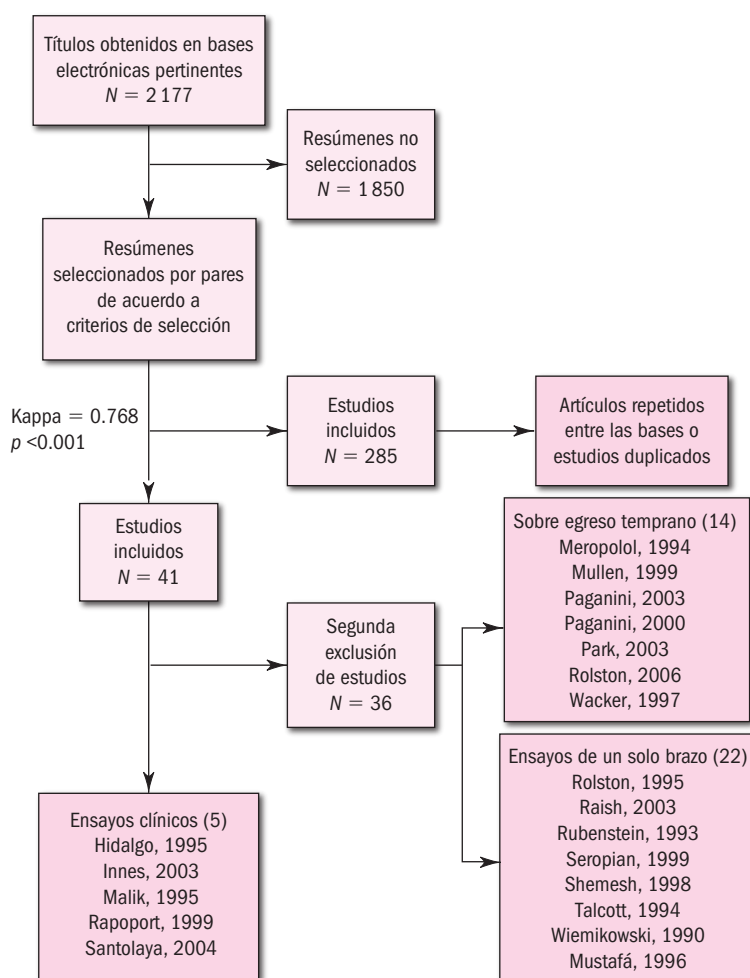


**Figura 8-4.** Ejemplo de la metodología de la selección de los estudios de un metaanálisis sobre el tratamiento antimicrobiano empírico ambulatorio en pacientes con cáncer, neutropenia y fiebre.

## Evaluación de la calidad metodológica y extracción de los datos

La calidad de los estudios originales que fueron elegidos para la revisión debe ser evaluada de manera uniforme por uno o varios investigadores, quienes deben utilizar un instrumento de evaluación que les permita, de forma estandarizada, extraer los datos de los estudios.

Se han publicado varios instrumentos con diferentes criterios de evaluación para calificar la calidad de los métodos de los estudios originales. Algunos de ellos son largos y complicados, como el de puntuación de calidad de Chalmers, en donde se califican los atributos del diseño de los ensayos clínicos controlados del estudio con base en 100 puntos, asignando diferente puntaje a las diferentes secciones del artículo. Algunos otros incluyen criterios mucho más sencillos, como los publicados por el grupo de McMaster, que sólo evalúa tres características:<sup>2</sup> 1) la aleatorización, 2) el cegamiento del estudio y 3) si existe una descripción completa de los sujetos que no terminaron el estudio.<sup>3</sup>

La colaboración Cochrane para las revisiones sistemáticas usa los criterios de Shultz,<sup>4</sup> que hacen especial referencia a la adecuada ocultación de la maniobra (*allocation concealment*) y establecen tres categorías: A = ensayos en donde se adopten las medidas adecuadas para ocultar la asignación (asignación al azar central; sobres cerrados, oscuros, numerados de forma consecutiva; u otra descripción que incluyera los ítems a favor de la ocultación); B = ensayos en los que los autores no informaron el enfoque de la ocultación de la asignación o informaron un enfoque que no corresponde a ninguna de las otras categorías, y C = ensayos en los que la ocultación fue inadecuada (como alternancia o referencia a los números de historia clínica o a las fechas de nacimiento).

Para los estudios observacionales se han diseñado instrumentos y criterios para evaluar la calidad metodológica. La más utilizada es la escala de New Castle/Ottawa para artículos de cohortes y casos y controles.<sup>5</sup>

Idealmente, las revisiones sistemáticas deben limitarse a estudios rigurosos desde el punto de vista metodológico, para lo cual varios grupos han propuesto criterios de fortaleza científica. Algunos ejemplos de esto son el grupo CONSORT para ensayos clínicos aleatorizados (ECA) ([www.consort-statement.org](http://www.consort-statement.org)), el STROBE para estudios observacionales (cohortes, casos y controles, y transversales) ([www.strobe-statement.org](http://www.strobe-statement.org)) o el grupo STARD para las revisiones de estudios de pruebas diagnósticas.<sup>6,7</sup>

Las diferencias en la metodología de los estudios originales quizá lleven a discrepancias importantes en las conclusiones, mismas que pueden tener implicaciones a nivel de salud pública. Es crucial que los estudios sean similares; es decir, homogéneos, a fin de que sea posible combinarlos.

## Resumen numérico de los estudios: metaanálisis

**Tamaño del efecto.** En los estudios de intervención, la unidad de medición es el tamaño del efecto (TE). El tamaño del efecto calcula la diferencia ( $\delta$ ) en promedio de los sucesos de interés del grupo experimental ( $\mu^E$ ) (expuestos o tratados) y del grupo control ( $\mu^C$ ) (no

expuestos) y se divide entre la desviación estándar del grupo control ( $\sigma$ ). El tamaño del efecto de cada estudio se convierte en una **nueva unidad de análisis**.

$$\delta = \frac{\mu^E - \mu^C}{\sigma}$$

Por ejemplo, suponga que en un estudio se comparan dos grupos de pacientes con espondilitis anquilosante. El primero recibe entrenamiento de fisioterapia en la institución donde es atendido por el médico especialista; en el segundo grupo se aplica un programa de fisioterapia domiciliaria. En ambos se utiliza un cuestionario que evalúa las actividades de la vida diaria. Después de un periodo de seguimiento, el promedio de puntuación del primer grupo es de 110, mientras que el del segundo grupo, de 100, con una desviación estándar de 20. El TE del grupo de fisioterapia en la institución es de  $(110/100)/20 = 0.5$ . Debe notarse que el expresar el TE en unidades de desviación estándar hace posible la comparación de las variables en los diferentes estudios. Algunos autores utilizan la desviación estándar combinada y no la desviación estándar del grupo control, pero la idea subyacente es la misma.

## Gráfico de árbol Forest Plot

Es un recurso que permite de forma esquemática ver en su totalidad el efecto del estudio. En este gráfico se puede observar el punto de estimación (efecto medio) de cada uno de los estudios y su intervalo de confianza (en general de 95%), el número de estudios revisados, la significancia estadística de cada uno de los estudios y el año en que fueron realizados; este gráfico permite evaluar situaciones que pudieran modificar el resultado de los ECA, como los posibles cambios a través del tiempo.

Los estudios observacionales y de exámenes diagnósticos también pueden ser resumidos en este tipo de gráficos.

En la figura 8-5 se pueden ver los resultados de un metaanálisis, que evalúa el tratamiento antimicrobiano empírico ambulatorio *vs.* el hospitalario en pacientes con cáncer, neutropenia y fiebre. La pregunta de esta revisión sistemática fue orientada a conocer la eficacia del tratamiento antimicrobiano empírico, en pacientes con cáncer que presentan neutropenia y fiebre. El gráfico muestra, de izquierda a derecha, en la primera columna estudios incluidos, seguidos del número de eventos (neutropenia) en cada grupo con el número de muestra. De forma gráfica en la parte derecha de la tabla se observa el efecto medio de cada estudio (cuadros sólidos en negro) y el efecto final (diamante negro al final). Las líneas a los lados (bigotes) representan el intervalo de confianza de 95%.

## Heterogeneidad, modelos de análisis e $I^2$

La validez de los metaanálisis depende de la calidad de los estudios a comparar y de lo similares que éstos sean; por esta razón, los metaanálisis deben garantizar que los estudios incluidos sean comparables. Como ya se señaló, una vez que se ha tomado la decisión de incluir a los estudios en un metaanálisis, es preciso evaluar qué tanta consistencia

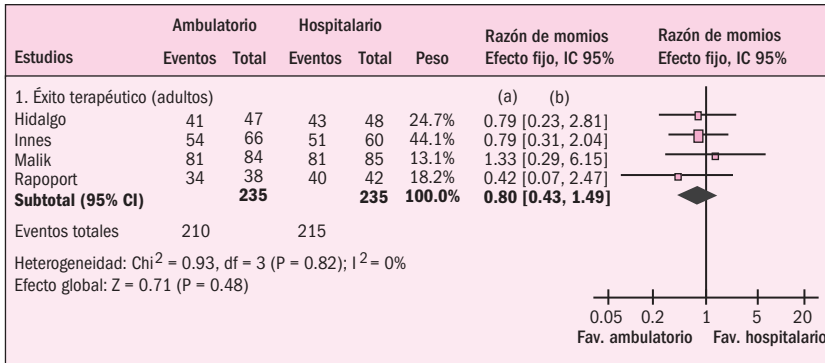


**Figura 8-5.** Gráfico de árbol de un metaanálisis, sobre la eficacia del tratamiento antimicrobiano empírico, administrado ambulatoria u hospitalariamente.

hubo entre los estudios individuales, tanto en la evaluación cualitativa (calidad de sus características y métodos) como en la estadística, es decir, medir la posibilidad de exceso de variabilidad entre los estudios; a esto se le ha llamado **prueba de heterogeneidad**. La consistencia de los estudios ofrece al lector la seguridad de que el resultado presentado sea confiable. La prueba de heterogeneidad utiliza la prueba de  $\chi^2$  para afirmar o descartar que los estudios incluidos sean estadísticamente parecidos. Cuando la prueba tiene un resultado estadísticamente significativo ( $p > 0.05$ ) se asume heterogeneidad, es decir que los resultados de los estudios no son similares. Cuando los estudios presentan heterogeneidad, es necesario analizarlos con el método de efectos aleatorios. Por otra parte, cuando los estudios son homogéneos, se utiliza el método de efectos fijos. En la figura 8-5 el estudio fue analizado con el método de efectos fijo, ya que la  $\chi^2$  fue no significativa (0.93). El análisis por efectos fijos asume que los estudios son consistentes entre sí.

Si bien la prueba de  $\chi^2$  permite asumir o no la homogeneidad de la prueba, no permite conocer la magnitud de la diferencia entre estudios. Para saber “cuán” heterogéneos son los estudios, se realiza la prueba de  $I^2$ , el cual se expresa como porcentaje, siendo el 0% el más homogéneo y el 100% el más heterogéneo.

Cuando los estudios son heterogéneos (heterogeneidad estadística), se asume que es debido a la diversidad clínica del fenómeno estudiado; en el ejemplo de la figura 8-6, la heterogeneidad muestra una  $p > 0.05$ , la  $I^2$  muestra “cuán” heterogéneo es el estudio; en este caso es de 79%, por lo que debe analizarse por efectos aleatorios (*random effects*).

### Gráfico de embudo (*funnel plot*)

Este gráfico es útil para evaluar el sesgo de publicación. En general, los estudios con una mayor cantidad de pacientes ( $n$  mayor), con resultados positivos, son más fáciles de publicar, que los estudios pequeños o negativos. El gráfico de embudo pretende ubicar a los estudios según su error estándar (es decir, su tamaño de muestra) y la relación con

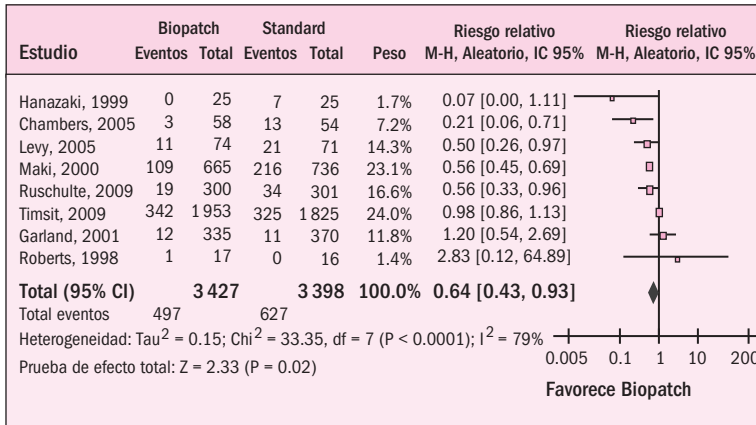


**Figura 8-6.** Ejemplo de un metaanálisis con heterogeneidad, analizado por efectos aleatorios.

el efecto final del metaanálisis. En un metaanálisis que no presente sesgo de publicación, se espera que el gráfico de embudo sea simétrico y que los estudios se ubiquen en ambos lados del efecto final del metaanálisis.

Como se aprecia en la figura 8-7, en el eje vertical se grafica el error estandarizado (SE, del inglés *standard error*). En el eje horizontal se ubica el efecto estimado del metaanálisis (línea punteada vertical); las líneas diagonales representan el intervalo de confianza del 95%, y los estudios están representados alrededor del efecto final.

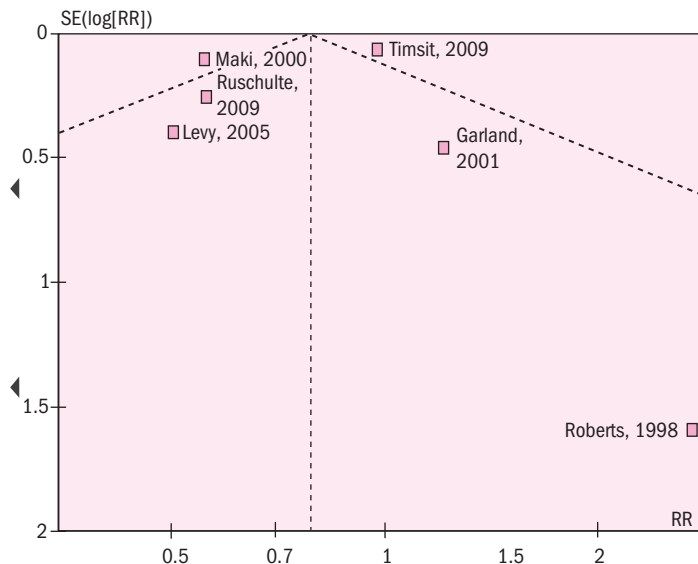


**Figura 8-7.** Embudo o *funnel plot* del metaanálisis de Biopatch.

Para aquellos lectores que deseen profundizar en los aspectos de análisis estadísticos de los metaanálisis, al final del capítulo se recomiendan varios textos relacionados.

## Colaboración Cochrane

La Colaboración Cochrane es una organización internacional, independiente y sin ánimo de lucro, establecida en el Reino Unido. Su principal objetivo es crear revisiones sistemáticas y metaanálisis, así como difundir esta información entre el personal de la salud y los pacientes. Para ello, promueve y apoya a los médicos y personal sanitario para la realización de las revisiones sistemáticas, otorgando apoyo en la traducción, realización y difusión de las mismas.

Las revisiones se publican periódicamente de manera electrónica, a través de The Cochrane Library (inglés) y en La Biblioteca Cochrane Plus (español), la cual es accesible de manera gratuita en México a través del portal de la Facultad de Medicina de la UNAM en la Biblioteca Médica Digital. En esta Biblioteca las revisiones se pueden obtener en español con un retraso sólo de un par de meses en relación con su original en inglés y en forma de texto completo, lo que le confiere una gran ventaja para la toma de decisiones.

La estructura de la Colaboración Cochrane es jerárquica; se basa en una administración central, que da apoyo a un grupo que desarrolla revisiones sistemáticas sobre una enfermedad o estado de salud (grupos de revisión en colaboración). Los grupos de revisión están distribuidos por todo el mundo y generalmente tienen su base en una Universidad u Hospital Universitario. En México existe un centro Cochrane que apoya a los investigadores que desean realizar revisiones sistemáticas.

## Evaluación de una revisión sistemática

Como se mencionó al inicio de este capítulo, no todas las revisiones sistemáticas y metaanálisis son realizados de forma estricta, por lo que es necesario contar con algunos estándares metodológicos para calificar a estos estudios. Actualmente están disponibles guías de lectura crítica para establecer la confiabilidad de estos estudios.

Las guías propuestas por el grupo de Medicina basada en la evidencia de la Universidad de McMaster son ampliamente utilizadas, y evalúan las siguientes características de los estudios:<sup>10,11</sup>

1. ¿La revisión aborda una pregunta clínica lógica y enfocada? En términos generales, en la pregunta de investigación es preciso buscar 3 o 4 componentes: si describe el tipo de paciente, el tipo de intervención (dosis, tratamiento, etc.), si se compara con un grupo control y si se hace referencia al desenlace. Tales características en la pregunta se conocen con el acrónimo de “PICO” de sus siglas en inglés (paciente, intervención, comparación y desenlace [*outcome*]). Evaluar a la pregunta de investigación de esta manera ayuda a valorar si la pregunta fue realizada para una aplicación clínica concreta.

2. Si se especificaron de forma clara los criterios de inclusión y exclusión de acuerdo con lo ya mencionado en este capítulo.
3. Si se determinó si la probabilidad de no inclusión de artículos relevantes puede ser un factor de sesgo en las revisiones. Esto se aprecia en los artículos revisando si la estrategia de revisión fue suficiente.
4. Si se evaluó la validez de los estudios incluidos, a través de escalas validadas para la extracción de los datos, y si incluyó uno o más revisores.
5. Si se especificaron de manera adecuada todos los procedimientos de forma tal que el estudio puede reproducirse por otros grupos.
6. Si los resultados fueron consistentes de estudio a estudio. Y si se determinó la heterogeneidad estadística en los estudios con alguna de las pruebas.

## Conclusiones

En este capítulo se han revisado algunas consideraciones del diseño y análisis estadístico que se utilizan para combinar los resultados de estudios originales. Esta técnica constituye un apoyo metodológico para la revisión estructurada de la literatura médica. Las revisiones sistemáticas y los metaanálisis son estudios de revisión, que por seguir el método científico son más confiables y con menos sesgos que las revisiones narrativas.

La información que se deriva de este tipo de estudios es de utilidad práctica para el clínico, los investigadores y tomadores de decisiones en salud. Las recomendaciones de este tipo de estudios son precisas y proveen información sobre la magnitud del efecto y el impacto directo del tratamiento en los ensayos clínicos, o de la probabilidad de los diversos factores de riesgos en diversas enfermedades provenientes de otros diseños, como casos y controles y cohortes.

La calidad de los estudios originales determina la certidumbre de los resultados en las revisiones sistemáticas y metaanálisis. Si la investigación original es de buena calidad, los resultados de las revisiones sistemáticas y metaanálisis serán en consecuencia de alta calidad y con un mayor grado de certidumbre.

## Respuesta del caso clínico

Con base en una revisión sistemática encontrada en la Biblioteca Cochrane Plus, sobre el tratamiento del síndrome de fibromialgia con diferentes esquemas de ejercicio que incluyen ejercicios aeróbicos, fortalecimiento muscular y de flexibilidad, en este metaanálisis publicado en el 2007 donde se revisó a 2 776 sujetos en 34 estudios incluidos, se encontró que el ejercicio aeróbico de intensidad moderada ejerce efectos positivos en 44% de los pacientes sobre el bienestar general, mejora la función física en 68% (fatiga), y mejora la depresión en 54%.<sup>12</sup> No se encontró diferencia en la disminución del dolor ni de los puntos en gatillo. Con base en este estudio es posible recomendar al paciente realizar ejercicios aeróbicos de intensidad moderada, además del tratamiento con fármacos, ya que es sabido que el tratamiento de la fibromialgia es multidisciplinario.

## Referencias

- Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA* 270:2093-2095. 1993.
- How to read systematic reviews and meta-analysis. *JAMA*. 2004.
- Jadad A, Moore R, Carroll D. Assessing the quality of reports of randomized controlled trials: is blinding necessary? *Controlled Clinical Trials* 17:1-72. 1996.
- Schulz K, Chalmers I, Hayes R, Altman D. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* 273:408-412. 1995.
- Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis. En [http://www.ohrica/programs/clinical\\_epidemiology/oxford.htm](http://www.ohrica/programs/clinical_epidemiology/oxford.htm)
- Moher D, Schulz K, Altman D, Group C. The CONSORT statement: revised recommendations for improvement the quality of parallel-group randomized trials. *Ann Int Med* 134:657-662. 2001.
- Bossuy P, Reitsma J, Bruns D *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 138:40-44. 2003.
- Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 358:1340-1342. 2001.
- Gøtzsche P, Nielsen M. Cochrane Database of Systematic Reviews. 2008.
- Guyatt G, Rennie D. *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. AMA Press, Chicago. 2002.
- Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analysis). *Br Med J* (Clinical research ed.) 315:672-675. 1997.
- Buch A, Barber K, Oberend T, Peloso P, Shachter C. Ejercicio para el tratamiento de la fibromialgia (Cochrane review). Cochrane Plus. 2008.

## Lecturas recomendadas y vínculos útiles para encontrar y usar revisiones sistemáticas

1. Egger M, Smith GD, Altman DG. *Systematic Reviews in health care: Meta-analysis in context*. 2a. ed. BMJ publishing group 2001.
2. Mulrow C, Cook D. *Systematic Reviews. Synthesis of best evidence for healthcare decisions*. American College of Physicians, 1998.
  - Cochrane: [www.cochrane-net.org](http://www.cochrane-net.org)
  - Cochrane Library Plus: [www3.interscience.wiley.com](http://www3.interscience.wiley.com)
  - Biblioteca Cochrane en español en la biblioteca médica digital.
  - <http://www.facmed.unam.mx/bmnd/>
  - PubMed: [www.pubmed.com](http://www.pubmed.com)
  - Embase: [www.embase.com](http://www.embase.com)
  - Clinical Trials: <http://clinicaltrials.gov/>





# Capítulo 9

## Clinimetría

*Niels H. Wachter*

### Introducción

Por muchos años en Medicina se consideró que obtener información de los pacientes era un arte. Algunas de las cosas que el médico pregunta a sus enfermos son completamente subjetivas. ¿Cómo hacer para que estos datos se conviertan en información objetiva? Los psicólogos y los investigadores de las ciencias sociales, que siempre han trabajado con ideas, opiniones y otros elementos subjetivos, desarrollaron una disciplina que llamaron **psicometría**, que les ayudó a mejorar la calidad de su información. En Medicina, no siempre es necesario desarrollar cuestionarios específicos, tenemos instrumentos de medición propios, como la historia clínica. Pero, como la información es la base de todas nuestras acciones, es importante revisar cómo sucede la obtención de datos de los pacientes y qué elementos pueden ocasionar errores en ese proceso; sobre todo, cómo hacer para obtener mejor información.

¿Bastará con leer un texto de Psicometría y otro de Estadística? Probablemente no (aunque no estaría de más). Los psicólogos y los matemáticos tienen sus propios puntos de vista; los primeros suelen buscar escalas monotónicas, de intervalos equidistantes, mientras que en Medicina, en la cual se cuenta con estándares de comparación, se eligen escalas con otros atributos (tensión arterial normal  $\leq 120/80$  mmHg, prehipertensión 121/81 a 139/89 mmHg e hipertensión  $\geq 140/90$  mmHg. Aquí los intervalos no son “equidistantes”). Además, en ausencia de un estándar de comparación, suelen evaluarse con cuestionarios con base en una cualidad que llaman **confiabilidad** (*reliability*), mientras que los médicos suelen comparar con un resultado objetivo. Para un matemático podrían ser más importantes el principio de reducción de la varianza y otros atributos, que para un médico no lo son, porque sabe qué límites debe considerar en las opciones de respuesta, y sabe con qué resultados se relacionan. Las Matemáticas y la Psicometría sí son aplicables en la Medicina, pero tenemos diferentes puntos de vista. En 1987, Alvan Feinstein publicó

su libro *Clinimetría*, que dedicó al estudio formal del proceso de medición en la clínica. Indudablemente la Clinimetría y la Psicometría tendrán muchas semejanzas y puntos de contacto, pero no son equivalentes.

## Caso clínico

---

Se presenta con el médico familiar un enfermo de 64 años de edad; él es diabético desde hace 23 años; se identificó historia de mal control metabólico; el paciente es además hipertenso y está con tratamiento para dicha enfermedad, dislipidemia y ácido acetilsalicílico. Un año antes sufrió un infarto al miocardio y se queja de disnea con esfuerzos moderados e intensos. Las pocas ocasiones en las que ha realizado esfuerzos intensos ha sufrido dolor opresivo en el centro del pecho, irradiado a la mandíbula, que desaparece aproximadamente dos minutos después de suspender el esfuerzo. El paciente trabaja como empleado en una empresa donde hace labores de oficina, por lo general está sentado frente a un escritorio. Recientemente consultó al oftalmólogo, de quien ha recibido tratamiento, primero con láser y después una vitrectomía por retinopatía diabética y el especialista señala que la visión de este paciente es de 80/400. El paciente, durante la consulta, se refiere a las ocasiones en las que ha sufrido opresión torácica y pregunta a su médico si él sería candidato a una pensión por invalidez por su enfermedad.

El caso no es del todo raro y el médico deberá tomar algunas decisiones con base en la información disponible: ¿es necesario enviar al paciente al cardiólogo? Esto depende de que se identifiquen algunas de sus molestias como angina de pecho y de que el médico las “clasifique”: angina estable o inestable, etcétera. Unas de ellas requieren el envío inmediato al hospital, otras a la consulta regular del cardiólogo. No basta con clasificar desde el punto de vista etiológico el de la angina, es necesario conocer su intensidad a fin de darse una idea de lo que puede ocurrir con el enfermo. También tendrán que considerarse las actividades habituales del paciente y lo que le requiere de esfuerzo su trabajo pues, además, debe dar respuesta a la interrogante principal del paciente: “¿debo solicitar una pensión por invalidez?”

---

El médico debe tomar diferentes decisiones frente al caso clínico que se le presenta. Algunas de ellas se llaman **decisiones explicatorias**, como es la necesidad de reconocer la molestia precordial y sus acompañantes, como angina de pecho (las decisiones explicatorias son las que se usan al realizar un diagnóstico o clasificar los datos), mientras que otras podrían llamarse **decisiones terapéuticas**: “¿qué debo hacer, enviarlo a un especialista, tratarlo o no hacer nada?” Es indiscutible la irresponsabilidad en que incurre un médico que no es capaz de tomar las decisiones apropiadas. Desde la antigüedad se reconoce que el trabajo del clínico es hacer el diagnóstico, establecer el pronóstico y ofrecer tratamiento.

La materia prima básica para tomar todas esas decisiones son los datos que se obtienen del paciente. Sólo si esos datos son correctos existe la oportunidad de tomar decisiones apropiadas.

Cada uno de esos datos se obtiene después de un proceso de medición. Algunos se obtienen con aparatos más o menos complejos (por ejemplo, los estudios de imagen, el

electrocardiograma, etc.) y otros, con aparatos más sencillos (la medición de la tensión arterial con un baumanómetro o la medición de la agudeza visual con una cartilla). Lo que comparten todos estos datos es que son “objetivos” (es decir que no dependen de la interpretación de quien hace la medición), pueden expresarse como un número que todo el mundo entiende (p. ej., 130/75 mmHg) y las más de las veces pueden guardarse en un registro para que se examinen después para conocer si el resultado (a menudo así sucede) es el mismo. A esta clase de información suele llamársele **datos duros** (porque son objetivos, se expresan cuantitativamente y pueden reanalizarse en el futuro). Sobre todo para los médicos más jóvenes, estos datos suelen ser muy atractivos y están rodeados de un “aura” de ser más científicos.

Otros datos corresponden con sensaciones y emociones, como es el caso de la angina de este paciente. Son subjetivos, no hay aparato para medirlos, no se expresan en números y suelen ser de naturaleza transitoria; de manera que no existe la posibilidad de analizarlos más tarde. A éstos suele llamárseles **datos blandos** y a pesar del aparente carácter derogatorio de su nombre, suelen ser los más importantes que se recogen del paciente, pues de esa información dependen el diagnóstico, el pronóstico y, con frecuencia, también el tratamiento (intente encontrar algún aparato que haga el diagnóstico de la angina de pecho, la depresión o la migraña, por nombrar sólo algunas enfermedades). ¿Cómo hacer para que estas sensaciones subjetivas se conviertan en datos para tomar decisiones apropiadas?

## Antecedentes

Por muchos años se consideró que la obtención de información del paciente es un “arte”. Así, la obtención de datos del paciente sólo estaría al alcance de unos cuantos “iniciados” y sólo algunos médicos serían capaces de obtener información apropiada de los pacientes. Como los datos blandos suelen ser los más importantes, sería imposible el estudio científico de la Medicina Clínica. Nada más alejado de la realidad; desde hace muchos años se han hecho esfuerzos para obtener información, aun datos blandos, con todo el rigor necesario para que sean sometidos al escrutinio científico.

John Haygarth (1740-1827) escribió un libro en el que detalló todos los eventos clínicos relacionados con más de 200 casos de artritis. Como resultado de esa publicación se estableció el uso de la cinchona (un polvo que se obtenía al moler la corteza del árbol de la quina) que aún puede encontrarse, aunque en forma sintética, en el tratamiento de la artritis; así, se desecharon numerosos tratamientos inútiles o perjudiciales. William Whittinger (1741-1799) describió en un libro poco más de 160 casos de hidropesía, que trató con el extracto de una florecilla (*digitalis* o dedalera) que sigue siendo (también con base en compuestos sintéticos) parte del tratamiento de la insuficiencia cardíaca.

Hacia la mitad del siglo XIX, Pierre Charles Alexandre Louis describió su *Methode Numerique* a través de un estudio que describía los resultados de la sangría en el tratamiento de las afecciones inflamatorias del tórax (la neumonía). Este médico tenía la costumbre de hacer descripciones minuciosas y detalladas de todos los eventos que ocurrían

a estos enfermos y del resultado de los tratamientos; también introdujo el estudio de las variables pronósticas, el análisis estratificado y el uso de promedios y tasas, que facilitaban la comprensión de la información que quería mostrar al lector (a diferencia de los trabajos de Haygarth y Whittering en los que había que leer todo el libro y los datos de cada paciente de manera individual). Con esos resultados se desechó una práctica que se usaba hacía más de 250 años en la Medicina, la sangría.

## Proceso de medición

Medir es identificar una cualidad y delimitar las categorías del resultado; como ya se señaló, la medición se puede hacer con los sentidos del médico o con ayuda de un aparato más o menos sofisticado, pero el proceso es el mismo en todos los casos:

- Identificar una cualidad.
- Comparar con el acervo de experiencias y conocimientos de médico.
- Clasificar el resultado.
- Enunciarlo.

Así, por sus características externas, el médico se puede percatar si una persona se ve igual o diferente de las otras (su color es diferente, identificar una cualidad). Acto seguido, compara lo que observó con lo que está acostumbrado a considerar como “normal” o lo que cree que correspondería en ese caso (compara con su acervo de conocimientos y experiencia, p. ej., “tiene un tinte ‘amarillo’ en la piel”), lo clasifica (“ictérico”, hasta qué punto o “normal”) y, finalmente, enuncia el resultado: “ictericia ++”.

Se trata de un proceso cognitivo, es decir, se lleva a cabo en la mente. Un ejemplo extraordinario es la clasificación de APGAR del recién nacido. Virginia Apgar, una anestesióloga del estado de Nueva York, después de años de experiencia señaló cinco características que determinan el pronóstico del recién nacido, inmediatamente después del parto. Ella decidió que esas eran las variables necesarias, después de asistir a numerosos partos. El único aparato que se requiere es el que se usa para contar la frecuencia cardíaca del recién nacido. Las demás variables: el tono muscular, la coloración de la piel, el esfuerzo respiratorio, etcétera, se perciben sin ningún auxiliar. Apgar delimitó las categorías de resultado. Si el recién nacido tiene una frecuencia cardíaca  $\geq 100$  latidos por minuto, se la considera normal y se califica con 2 puntos; si está por debajo de este valor, se le asigna un punto y si no es detectable, se califica con cero. Cada una de las variables se califica y el resultado de cada una se suma en una puntuación final que arroja resultados entre 0 y 10 puntos (figura 9-1).

| Variable            | 0       | 1                      | 2               |
|---------------------|---------|------------------------|-----------------|
| Color               | Pálido  | Cianótico              | Rosado          |
| Frecuencia cardíaca | Ausente | <100 × '               | <100 × '        |
| Respiración         | Ausente | Irregular, superficial | Llanto enérgico |
| Reflejos            | No      | Gesticula              | Enérgico        |
| Tono muscular       | No      | Flexión                | Adecuado        |

**Figura 9-1.** Escala de Apgar para evaluación del estado del recién nacido.

## Componentes del proceso de medición

Todo proceso de medición puede fallar, dando lugar a la variabilidad. Por eso es importante disecar sus componentes para conocer cómo contribuye cada uno de ellos a la variabilidad. En todo proceso de medición intervienen los siguientes componentes (figura 9-2):

- Variable.
- Instrumento de medición.
- Método de medición.
- Sujeto de la medición.
- Individuo que hace las mediciones (el observador).

A continuación se describe cada una de ellas.

### Variable

Por el solo hecho de notar una cualidad, ésta puede tomar al menos dos valores (presente o ausente) y por esa razón a esas cualidades que medimos se les llama: variables. Algunas

- **Variable:** retumbo diastólico vs. escape aórtico
- **Instrumento de medición:** estetoscopio barato reparado con una radiografía vs. Master Cardiology III
- **Método:** 4 focos vs. 4 focos + uso del foco accesorio aórtico + maniobras especiales (hacia adelante, Valsalva, etc.)
- **Sujeto de la medición:** delgado u obeso
- **Observador:** cardiólogo experimentado vs. residente

**Figura 9-2.** Ejemplo de la contribución de los componentes del proceso de medición a su resultado.

variables son más fáciles de medir que otras. Por ejemplo, el soplo de la insuficiencia mitral (retumbo diastólico) suele ser más sonoro y fácil de descubrir que un soplo de insuficiencia de la válvula aórtica (escape aórtico).

Existen diferentes tipos de variables. Las variables **cualitativas** son de dos tipos: cuando sólo nos percatamos de su presencia se les llama variables **nominales** (p. ej., nacionalidad, hombre o mujer, etc.). Cuando es posible observar diferente intensidad en algunos sujetos, pero se desconoce de qué tamaño es la diferencia entre uno y otro, las variables resultantes reciben el nombre de **ordinales**, es decir, sólo se logra identificar qué lugar ocupa un individuo en el grupo (orden). Los pacientes con ictericia se clasifican según su intensidad en +, ++, +++, etcétera. Aunque es evidente que + es menor que ++ y, en consecuencia, también menor que +++, no puede afirmarse que ++ sea el doble de +.

Cuando sí se ha determinado la distancia entre una y otra intensidades, las variables son **cuantitativas** y también se clasifican en dos tipos: **discretas**, que son aquellas en las que se sabe la distancia entre dos valores pero no es posible fraccionarlas (p. ej., cuántos hijos tiene una familia 0, 1, 2, etc., pero no hay puntos intermedios). Además están las variables **continuas**, aquellas en las que sí es factible identificar fracciones entre uno y otro valores, dependiendo de la precisión del instrumento (p. ej., una persona puede tener un peso de 52 kg o 52.2 kg, 52.250 kg, etc.) (figura 9-3).

Otros tipos de variables son, por ejemplo, un **indicador**, mismo que es una variable que expresa una sola dimensión (**unidimensional**); un **índice** es una variable que se compone de muchas otras y es **multidimensional** (como la escala de Apgar, que suma los puntajes de cinco variables en una sola expresión). Un **proxyindicador** es una medición relacionada con la variable de interés, por lo que resulta indirecta (como la medición de la tensión arterial con esfigmomanómetro, en lugar de colocar una cánula intraarterial para medir directamente la presión arterial). Es importante distinguirlos porque el pro-

Tipos de variables

| Variable     |          | Características                                                                                        |
|--------------|----------|--------------------------------------------------------------------------------------------------------|
| Cualitativa  | Nominal  | Clasificación exhaustiva de categorías mutuamente excluyentes                                          |
|              | Ordinal  | Se distinguen intensidades, distancia no conocida (aplican algunas características de los números >,<) |
| Cuantitativa | Discreta | Distingue intensidades, intervalos sin fracciones (aplican todas las características de los números)   |
|              | Continua | Distingue intensidades, se puede fraccionar (aplican todas las características de los números)         |

Figura 9-3. Clasificación y características de las variables.

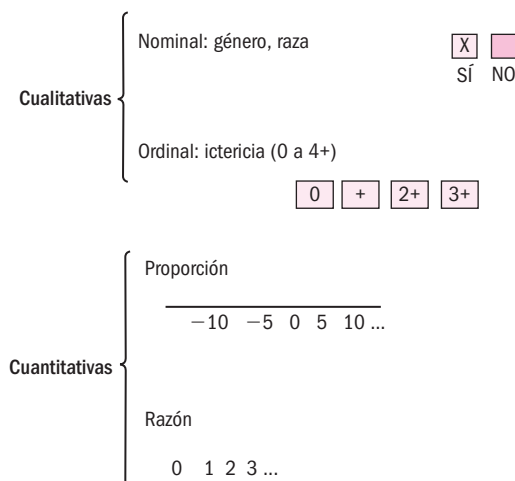


**Figura 9-4.** Escalas de medición.

xyindicador no será útil en algunas circunstancias (la medición de la tensión arterial es inexacta en los pacientes con hipotensión grave y en ellos debe medirse la presión arterial directamente).

Las variables se expresan en escalas de medición, que también se clasifican en **cualitativas** (**nominales** y **ordinales**) y **cuantitativas**. Hay dos tipos de escalas cuantitativas: escalas de **razón** (el cero identifica la ausencia de esa cualidad; es decir, el cero es verdadero), como en la escala Kelvin, donde el cero significa la total ausencia de energía o calor; mientras que en las de **proporción** el cero es convencional. Es decir, se acordó que a un cierto valor se le llamaría cero y que a los valores por debajo de éste, se les asignará un signo negativo (como la escala Celsius de los grados centígrados) (figura 9-4).

Las variables también pueden clasificarse en dependientes (efecto) e independientes (causas).

## Instrumento de medición

La calidad del instrumento es determinante del resultado y la calidad de las mediciones (compare cómo se oyen los ruidos cardiacos con un estetoscopio corriente y cuya membrana se reparó con un pedazo de radiografía, contra lo que ocurriría si usa un estetoscopio especializado para cardiólogo). Sin embargo, quizá lo más importante de un instrumento de medición es la relevancia de la variable que mide, en relación con lo que al médico le interesa saber; por ejemplo, un estudio de resonancia magnética nuclear, con todo su atractivo, no sirve para diagnosticar una migraña, que se caracteriza por dolor que tiene un patrón específico de presentación (hemocránea o universal, recurrente, etc.) y fenómenos acompañantes (náusea, acúfenos, fosfenos, etc.). Además, siempre será mejor que el instrumento sea accesible, de bajo costo y de fácil operación.

Es importante distinguir los diferentes tipos de instrumentos de medición:

- a) *Clasificación*. Un instrumento de clasificación se diseña con la intención de distinguir sujetos con base en sus características. Entre éstos se incluyen aquellos que se usan para conocer el estado socioeconómico o la calidad de vida de las personas y muchos más. Como se diseñan para formar el mayor número posible de subgrupos, los elementos que lo forman (preguntas o ítems) son aquellos que muestran la máxima varianza.
- b) *Predicción*. Un instrumento de predicción se diseña con la intención de conocer el estado futuro de una persona, con base en sus características actuales. A este tipo de instrumento corresponden las escalas de diagnóstico (como los criterios de Jones para la fiebre reumática) y los de pronóstico (como la escala APACHE II de los pacientes de la terapia intensiva). El requisito más importante es que los reactivos (*ítems*) estén fuertemente asociados con el desenlace que se quiere predecir y no se pone tanta atención a la varianza.
- c) *Evaluación*. Son instrumentos diseñados para identificar cambios (como la medición de la glucosa en la diabetes) en el tiempo y, por esa razón, los reactivos que los forman deben ser aquellos con la varianza más pequeña, pues, si mostraran tanta variabilidad como los instrumentos de clasificación, una diferencia pequeña pero clínicamente importante sería indetectable y no serían sensibles a los cambios en el tiempo.

Es evidente que los instrumentos de medición no son intercambiables, pues se diseñan con características opuestas (p. ej., mayor o menor varianza). Antes de aplicar un instrumento para un propósito diferente de aquel con el que se diseñó, debe probarse que funciona adecuadamente para este nuevo propósito.

## Método de medición

De poco sirve un instrumento costoso si no se usa de manera correcta. El estetoscopio especializado y costoso no ayudará a detectar el escape aórtico si el médico no explora el foco accesorio aórtico (además de cuatro focos tradicionales: pulmonar, aórtico, mitral y tricuspídeo) y si no hace maniobras especiales durante el examen (maniobra de Valsalva, espiración forzada, etc.).

Las instrucciones para usar el instrumento de medición suelen llamarse **descripción operativa**, es decir, la descripción de todos los procedimientos necesarios para obtener la información: identificación, clasificación, registro, captura, conservación de datos o muestras, transporte, además del proceso mismo de la medición. Esta descripción deberá incluir, cuando los haya, factores que causan interferencia (por ejemplo, ruido ambiental si quiere auscultar a un paciente, luz artificial cuando trata de descubrir si el paciente tiene ictericia), si hay controversias y cómo se resuelven éstas.

Siempre se obtienen mejores resultados cuando los criterios de juicio son explícitos (como en el caso de la frecuencia cardiaca en la escala de Apgar) que cuando se usan criterios implícitos (cuando no existen límites claramente demarcados, como cuando se cata-loga a la ictericia como +, ++ o +++).



## Sujeto de la medición

Las características del sujeto de la medición afectan los resultados: es más fácil escuchar los ruidos cardíacos en un sujeto delgado (el aire de los pulmones y la grasa del cuerpo dificultan el paso del sonido desde su origen en el corazón) que en un sujeto obeso. Una persona con un mejor nivel cultural o de más inteligencia puede responder mejor y más claramente las preguntas, y es más fácil detectar la palidez de la anemia en una persona rubia que en una más morena. Así, la edad, las características antropométricas y la cultura, entre otros datos, pueden facilitar o dificultar la obtención de la información.

## Observador

El último componente del proceso de medición es el sujeto que obtiene las mediciones, si este individuo es menos sensible a las diferencias culturales o del idioma; si tiene algún déficit sensitivo (sordera parcial, entre otros) obtendrá mediciones de menor calidad. Si está mejor entrenado y tiene más experiencia, podrá obtener mejores resultados.

## Atributos de una medición

Toda medición tiene dos atributos independientes: consistencia y validez. Se pueden tener ambos, cualquiera de ellos o ninguno. La primera y posiblemente la más importante es la consistencia, pues no tiene mucho sentido buscar la validez de una medición cuando el resultado es inconsistente.

## Consistencia

La consistencia tiene muchos sinónimos: repetibilidad, reproducibilidad, confiabilidad, etcétera. Una medición es consistente cuando, al repetirla en las mismas condiciones, arroja el mismo resultado o uno muy parecido.

¿Por qué es tan importante la consistencia? Lo es porque al usar las mismas unidades de medición es posible hacer referencia a una cualidad y siempre saber de qué se trata y en qué magnitud está presente. Las unidades de medición suelen ser arbitrarias. Todo el mundo sabe qué es un metro. Esta unidad de medida sustituyó al pie, al codo y a la vara en el Sistema Internacional de Unidades. Si alguien refiere el resultado de una medición en metros (centímetros o kilómetros), todo el mundo entiende qué fue lo que se encontró. El pie se refería a la longitud del pie del rey (imagine el lector la talla del calzado de Carlo Magno,  $\approx 30$  cm), pero otros reyes seguramente usaban calzado de talla diferente. La primera definición de un metro era una fracción de la circunferencia de la Tierra, que puede variar con los accidentes geográficos y otras condiciones; luego, se hizo un patrón de platino de esa longitud y se guardó en París. Pero, como los cambios de temperatura afectan la longitud de los metales y se requería algo más “consistente”, éste se cambió por la amplitud de la onda de luz de un gas raro. En la actualidad se define al metro como la distancia que recorre la luz en el vacío en  $1/299\,792\,458$  partes de un segundo (17a. Conferencia General de Pesas y Medidas). La medida es arbitraria, pero como es consistente todo el

mundo la usa. ¡Pida a dos personas que le señalen cuánto mide su escritorio en codos y compare ese resultado con una medición en centímetros!

Ante un paciente con insuficiencia cardíaca, el médico puede expresar el resultado con base en la clasificación de la Asociación de Cardiólogos de Nueva York (NYHA) y si se señala que ésta es clase 1, 2, 3 o 4, hace la distinción entre el paciente asintomático, del que está confinado a la cama y tiene síntomas en reposo. Estas “escalas” identifican una cualidad y delimitan los valores del resultado, de acuerdo con un sistema preestablecido. Esto las hace “consistentes” y útiles para el trabajo cotidiano, aunque la disnea, el síntoma más importante de la insuficiencia cardíaca, es una sensación subjetiva. Cuando se expresa en un sistema consistente, se convierte en una variable más “dura” y objetiva. ¿Cuál de las diferentes clases de la NYHA de insuficiencia cardíaca consideraría leve, moderada o grave?

La consistencia es un atributo que puede medirse y rara vez se alcanza de manera absoluta. Para evaluar la consistencia de una medición, se requiere que se hagan mediciones, al menos por duplicado, en una muestra de sujetos que contenga todos los grados de anormalidad que habrían de evaluarse (de otra manera sólo se sabría cómo funciona la medición en un porcentaje de los posibles valores), que las mediciones se hagan de manera independiente y “a ciegas”; es decir, que quien hace la segunda medición no conozca el resultado de la primera (y así debe expresar lo que vio y no lo que esperaba encontrar), que se usen criterios de interpretación claros, explícitos y aceptables para todos los evaluadores (a diferencia de criterios “implícitos” que no se definen; por ejemplo: mucho y poco, más o menos, leve, moderado y grave) y que se establezcan categorías de respuesta que sean mutuamente excluyentes (quien corresponde a una categoría no puede ser parte de otra).

Se reconocen dos tipos o clases de consistencia:

- **Consistencia intraensayo o intraobservador.** Se refiere a la repetibilidad de las mediciones que hace un solo individuo o que se hacen con un solo método.
- **Consistencia interensayo o interobservadores.** Se refiere a que dos o más individuos o que con dos procedimientos diferentes, se obtengan iguales resultados. La consistencia interobservadores siempre es menor que la consistencia intraobservadores, pues siempre que se evalúa entre sujetos cada uno de ellos aporta su propia consistencia.

## Validez

A la validez también se le llama certeza, conformidad o precisión. Se refiere a qué tanto concuerda el resultado de la medición con el valor real. Cuando existe un patrón o estándar de oro, se refiere a qué tanto concuerda ese estándar de oro con el resultado de la medición (p. ej., si concuerda el resultado de una prueba diagnóstica con el de una biopsia o un cultivo). El concepto de “validez” es más amplio y también se refiere a qué tan bien y qué tan completamente refleja la verdad esa medición.

## Tipos de validez

¿Qué tanto concuerda el resultado de una medición con el valor real?; esto depende de diferentes juicios y por esta razón se reconocen diferentes tipos de validez.

- **Validez interna y externa.** Toda la información recibida se obtiene del estudio de un número limitado de individuos (muestras), porque sería imposible estudiar a toda la población. Se dice que los resultados de un estudio tienen validez interna, cuando los datos se obtuvieron de acuerdo con el plan de trabajo preestablecido y están libres de sesgo; es decir, son creíbles. Por otro lado, se dice que estos mismos datos tienen validez externa, cuando son aplicables a otra población. Por ejemplo, suponga que desea conocer la talla promedio de las personas del país; para esto decide medir a los estudiantes de su universidad. Establece un programa que le permite medir la estatura de cada estudiante con mucha precisión y sin fallas. En ese caso, los datos tendrían validez interna pero, ¿cree que estos resultados son aplicables a la población de todo el país, o a las personas de mayor o menor edad, o a otros grupos (no estudiantes), o a las personas de las diferentes regiones geográficas? La diferencia no es ociosa. Las personas estudiadas ya son parte del pasado y lo más probable es que el investigador no las volverá a ver.

Usted quiere los datos de un estudio, para estimar qué ocurriría con los nuevos pacientes y por eso es necesario que el estudio tenga validez interna y externa. Si el estudio no tiene validez interna (p. ej., el instrumento sobrestima la talla de cada sujeto), los resultados son inservibles. Pero, si la talla se midió de manera precisa, los resultados aún son útiles para otros estudiantes universitarios, aunque no sean aplicables a todo el país. Una falla en la validez interna hace que los resultados sean inútiles; una falla en la validez externa limita la utilidad de los resultados, es decir, no podrán extrapolarse a una población con características semejantes.

- **Validez de apariencia.** En inglés se le ha llamado *face validity* y corresponde con un juicio subjetivo que señala en qué medida el resultado refleja apropiadamente la cualidad que se intenta medir: por ejemplo, ¿cuál de los siguientes exámenes tiene mejor validez de apariencia para evaluar la capacidad de un médico, el examen de opción múltiple o el examen práctico con pacientes?
- **Validez de contenido.** Se trata de un juicio que corresponde a cuán completamente se midió la cualidad que se intenta medir. Por ejemplo, una medición de la calidad de vida tiene más validez de contenido si pregunta aspectos relacionados con el dolor, la funcionalidad (movilidad, capacidad para cumplir el rol social y familiar), las emociones (sensación de bienestar, ansiedad y depresión), etcétera, que otra que se limita a examinar las emociones.
- **Validez de expresión.** Se refiere a que se usaron las unidades de medición apropiadas. Por ejemplo, el dolor no se puede expresar en unidades de longitud (centímetros). Si se usó una escala visual análoga de 10 cm para medir la intensidad del dolor, el resultado en centímetros sólo sirve como escala ordinal (para señalar quién tiene más dolor y quién tiene menos dolor), y un resultado de 2 cm no es la mitad de otro de 4 cm.

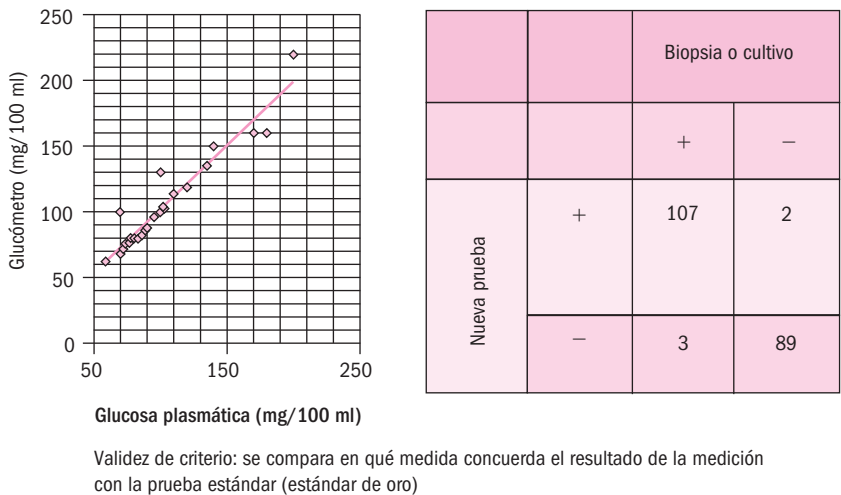
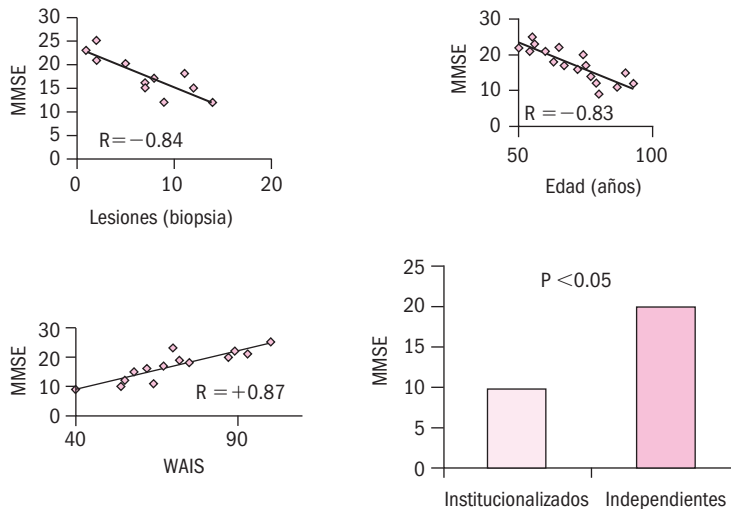


Figura 9-5. Validez de criterio.

- **Validez de criterio.** Quizá se trata de la mejor forma de evaluar la validez, pero sólo puede aplicarse cuando existe un estándar de comparación (estándar de oro). Consiste en evaluar qué tanto concuerda el resultado de la medición con el valor verdadero. Así, usted podría evaluar la validez de criterio de un glucómetro, aplicando la medición en el contenido de una gradilla de tubos de ensayo que contengan concentraciones conocidas de glucosa o contra el resultado de la medición de la glucosa plasmática (medida en el laboratorio) tomada al mismo tiempo en esa misma persona. Una prueba de diagnóstico podría compararse con el resultado de una biopsia o de un cultivo, y una escala que evalúa pronóstico se compararía con el desenlace final de la enfermedad, después de un periodo de observación apropiado (figura 9-5).
- **Validez de “constructo”.** Un “constructo” es un concepto o una idea; algunos ejemplos son la demencia y la calidad de vida. No existe un estándar de calidad de vida y ningún aparato para medirla y cuando esto sucede, debe evaluarse qué tanto concuerda el resultado de la medición con lo que esperaría suceda en diferentes circunstancias si ese concepto (constructo) resulta cierto.

Esto se consigue a través de dos estrategias: **convergencia** y **discriminación**. Por ejemplo, cuando se diseñó una breve escala de cinco preguntas para indagar, junto a la cama del enfermo, si el paciente tiene demencia (pérdida de las funciones mentales superiores), llamada *Minimental State Examination* (MMSE, por sus siglas en inglés), se correlacionaron (**convergencia**) los resultados que se obtuvieron con esta prueba en un grupo de personas con los obtenidos en una prueba de inteligencia (*Weschler Adult Intelligence Scale*, WAIS) y en aquellos que tenían biopsia del cerebro, con el número de lesiones que suelen aparecer en la enfermedad de Alzheimer (una forma de demencia, haces neurofibrilares y placas de



Validez de constructo: no se espera que los resultados concuerden, se espera que muestren la tendencia esperada (convergencia) o las diferencias esperadas (discriminen)

**Figura 9-6.** Validez de constructo.

amiloide) y finalmente con la edad de las personas. Así, aunque algunas personas sean más inteligentes que otras, es de esperarse que quienes tienen puntajes más bajos en el cuestionario MMSE, también muestren menores puntajes en la prueba de inteligencia (una correlación  $r$ , con signo  $+$ ), quienes tienen puntajes más bajos en el cuestionario MMSE, también muestren más lesiones por campo en las biopsias (una correlación  $r$ , con signo  $-$ ) y que las personas de más edad (en quienes es más frecuente la demencia), también muestren menores puntajes en el cuestionario MMSE (una correlación  $r$ , con signo  $-$ ). Por otro lado, es más probable que, después de un tiempo de observación, las personas con puntajes más bajos en la prueba (MMSE) tengan que vivir institucionalizadas (en un asilo de ancianos) y que las que mostraron valores más altos aún vivan de manera independiente (**discriminación**). Ninguna de estas mediciones es el estándar de oro (inteligencia, lesiones en la biopsia, edad, vivir institucionalizado), pero en conjunto señalan que el resultado de la prueba (MMSE) concuerda con lo que esperaríamos encontrar en una persona con demencia y, en consecuencia, nos permiten señalar si el resultado de la prueba es válido (figura 9-6).

### Confiabilidad (*reliability*)

Un tercer atributo suele tratarse con más profundidad en los textos de Psicología o en el área de la investigación en educación. Con frecuencia causa confusión, pues tanto la consistencia como la validez forman parte de la evaluación de la confiabilidad. Aquí me habré de referir a un tercer tipo de “confiabilidad”.

| Paciente          | Medición 1                                       | Medición 2           | Promedio individuos                    |
|-------------------|--------------------------------------------------|----------------------|----------------------------------------|
| 1                 | 104                                              | 99.9                 | 104                                    |
| 2                 | 82                                               | 72                   | 77                                     |
| 3                 | 90                                               | 86                   | 88                                     |
| 4                 | 83                                               | 77                   | 80                                     |
| 5                 | 136                                              | 127                  | 131.5                                  |
| 6                 | 25                                               | 31                   | 28                                     |
| 7                 | 86                                               | 100                  | 93                                     |
| 8                 | 100                                              | 98                   | 99                                     |
| 9                 | 169                                              | 150                  | 159.5                                  |
| 10                | 95                                               | 70                   | 82.5                                   |
| 11                | 142                                              | 125                  | 133.5                                  |
| 12                | 96                                               | 109                  | 102.5                                  |
| 13                | 93                                               | 124                  | 108.5                                  |
| 14                | 200                                              | 194                  | 197                                    |
| 15                | 97                                               | 88                   | 92.5                                   |
| 16                | 70                                               | 77                   | 73.5                                   |
| 17                | 132                                              | 144                  | 138                                    |
| 18                | 107                                              | 97                   | 102                                    |
| Promedio columnas | Promedio 1<br>105.94                             | Promedio 2<br>104.05 | Varianza individuos<br>$\sum_i = 1384$ |
|                   | Varianza mediciones 1 y 2<br>$\sigma^2_M = 1.77$ |                      |                                        |

#### Confiabilidad (*reliability*)

El promedio de dos mediciones de 18 pacientes es  $X = 105.02$  y la varianza total es  $\sigma^2_T = 1432.26$

La varianza total se puede descomponer en tres elementos = varianza de individuos, varianza de las mediciones 1 y 2 y una tercera llamada varianza de error

$$\text{La } \sigma^2_T = \sigma^2_i + \sigma^2_M + \sigma^2_e$$

$$(1432.26 = 1384 + 1.77 + \sigma^2_e)$$

$$\text{Donde } \sigma^2_e = 46.49 \quad (\sigma^2_e = \sigma^2_T - (\sigma^2_i + \sigma^2_M))$$

$$\text{La confiabilidad (reliability)} = \sigma^2_i / \sigma^2_T = 0.96$$

Esto significa que 96% de la variabilidad se explica por las diferencias que hay entre un individuo y otro

**Figura 9-7.** Confiabilidad (*reliability*).

En todas las mediciones es esperable algún grado de variabilidad. Si usted hace la misma medición en dos ocasiones a un grupo de 10 sujetos, obtendrá 20 valores de resultado (los 10 de la primera medición y los 10 de la segunda medición). La variabilidad se puede cuantificar por medio de la varianza o la desviación estándar. La variabilidad (varianza) total tiene al menos tres componentes: *a*) las diferencias entre un sujeto y otro; *b*) las diferencias entre las dos mediciones de un mismo sujeto, y *c*) una tercera fuente de variabilidad que no es posible explicar y que por esta razón recibe el nombre de **error**.

La razón fundamental por la que se efectúa cualquier medición, es porque se desea distinguir a un individuo de los demás. Esto sólo sucede cuando los resultados de la medición difieren más de un individuo a otro que entre dos mediciones de un mismo individuo. Estas mediciones de confiabilidad suelen expresarse con coeficientes que resultan de dividir la variabilidad entre los individuos (varianza entre sujetos) entre la variabilidad total de los resultados (la varianza total, que es la suma de todas las fuentes de variación). El resultado es siempre un valor  $\geq 0$  y  $\leq 1.0$ . Si el resultado de esta medición es un valor alto (p. ej.,  $\geq 0.7$ ) la mayor parte de la variabilidad se debe a diferencias de un sujeto a otro y no a las otras fuentes de variación y entonces se dice que la medición es confiable; puede distinguir entre uno y otro sujeto (figura 9-7).

## Variabilidad y error en la medición

Cuando ocurren fallas en la consistencia o en la validez, se dice que ocurrió **error de medición**. Si la falla es en la consistencia se le llama **error aleatorio** o simplemente error.

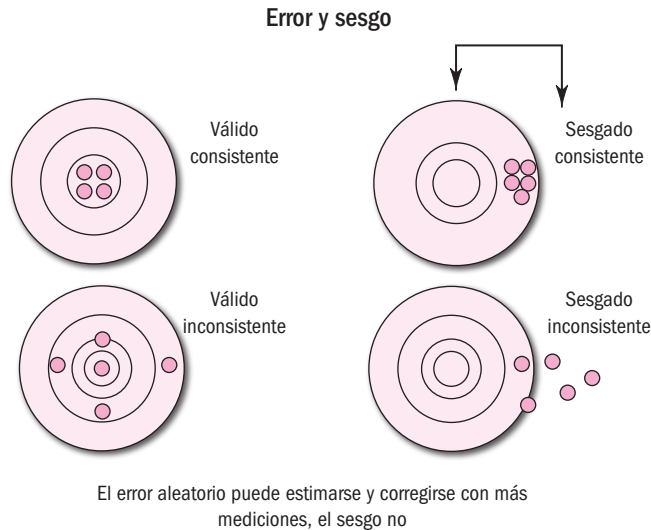


**Figura 9-8.** Error aleatorio y sesgo.

Cuando la falla ocurre en la validez, se dice que ocurrió **error sistemático** o **sesgo**. Estos dos tipos de error pueden ocurrir solos o en combinación en cualquier medición (recuerde que la consistencia y la validez son atributos independientes) (figura 9-8). Esta variabilidad puede ser consecuencia del mismo sujeto de la medición (los hay altos, bajos, delgados y obesos, etc.), del proceso de la medición y de la forma en que el observador expresa los resultados (variabilidad de expresión; por ejemplo, ictericia leve o +).

### Error aleatorio

Cuando los resultados de mediciones repetidas arrojan indistintamente valores por arriba y por debajo del valor real, se dice que ocurrió **error aleatorio** o simplemente **error**. Esta forma de error ocurre sobre todo asociado a la variabilidad biológica, inatención del observador o porque se usaron criterios implícitos para clasificar los resultados.

Para que el error se considere como aleatorio, su promedio debería ser cero o un valor muy cercano a cero, y no debe correlacionarse con la magnitud del resultado de la medición. Es decir, la diferencia entre el valor real y el observado suele mantenerse a través de todos los valores de  $\chi$ . La consecuencia más importante del error aleatorio es que oscurece relaciones, subestima correlaciones y reduce la magnitud de las medidas de asociación (figura 9-9) (la razón de momios o el riesgo relativo). En un estudio de factores de riesgo, el error en la medición de la exposición causaría que una proporción de los que se supone estaban expuestos, no lo están y otra proporción de los que se supone no estaban expuestos, sí lo estarían.

Así, por ejemplo, no todas las personas saben con precisión si sus padres son hipertensos. Suelen informarse proporciones similares de padres hipertensos, que no lo

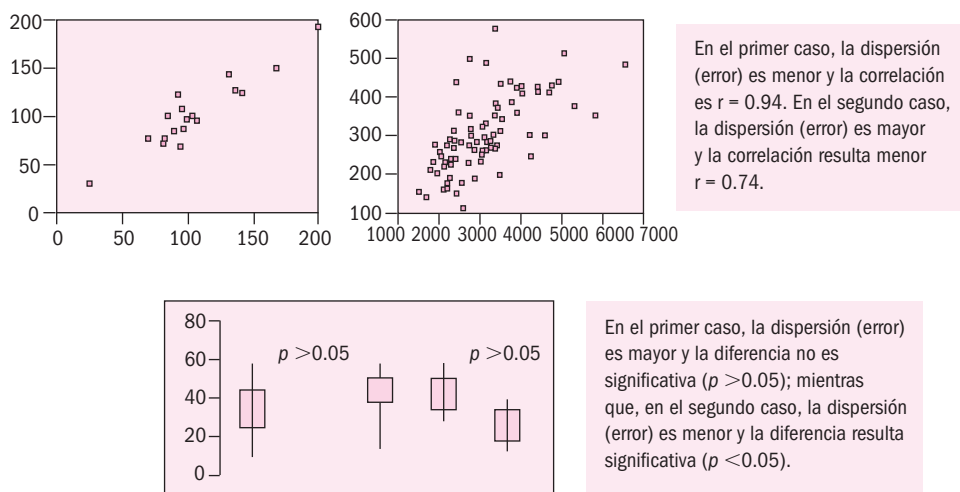


**Figura 9-9.** Consecuencias del error aleatorio.

son, que de padres que no son hipertensos, cuando en realidad sí lo eran. Al tratar de relacionar esta “exposición” con alguna consecuencia (como hipertensión en los hijos), la razón de momios aparecería más pequeña de lo que en realidad es. Si la diferencia es estadísticamente significativa, esto no importaría tanto (pues se llegaría a la misma conclusión), pero se subestima la magnitud del riesgo y en ocasiones este error ocasiona que la diferencia que debía ser significativa, aparezca como nula ( $p > 0.05$ ). Si la magnitud del error no es grande, su presencia es aceptable y no afecta mayormente los resultados finales. Pero si el tamaño del error es mayor, la situación aún puede resolverse con algunas estrategias, como aumentar el tamaño de la muestra, hacer mediciones repetidas y usar el promedio de ellas en lugar de los datos originales.

En términos generales, es esperable menor consistencia cuando existen diferencias pequeñas o menor grado de anormalidad entre las categorías de resultado y cuando los criterios para distinguir quién corresponde a cada una son implícitos. Es de esperarse que exista mayor concordancia cuando participan menos observadores, cuando hay más sujetos con resultados anormales en el grupo a medir, cuando hay menos opciones de respuesta, cuando los observadores recibieron más y mejor entrenamiento, cuando se establecieron criterios explícitos para definir quién corresponde a cada categoría de resultado y cuando se establecieron anticipadamente y por consenso las definiciones, terminología, criterios y reglas de decisión. Poner atención a todos estos aspectos puede mejorar la consistencia de una medición.

## Error sistemático o sesgo

Cuando la medición arroja valores que consistentemente se encuentran por arriba o por debajo del valor real, se dice que ocurrió **error sistemático** o **sesgo** (también se le define



como: prejuicio, opinión antes de conocer los datos, tendencia, inclinación, distorsión, divergente, apartado y torcido). Por definición, el **sesgo** significa que el promedio de la diferencia entre el valor observado y el real es  $\neq$  de 0 y puede tener una correlación  $\neq$  0 con el valor original, es decir, suele ser mayor o menor de acuerdo con el valor de  $\chi$ . El sesgo puede ocurrir en cualquier parte del proceso de inferencia: al seleccionar las referencias, en el diseño del estudio, la selección de los sujetos de investigación, en la aplicación de la maniobra experimental, por confusión, en el análisis e interpretación de los datos hasta en la decisión de publicar o no los resultados, y no sólo durante la medición (que se ha revisado a mayor profundidad en este capítulo). El sesgo tiene como consecuencia que conduce a resultados sistemáticamente apartados de la realidad.

En medición (sesgo de información), se asocia con las condiciones de la medición (p. ej., al tratar de detectar ictericia en una habitación iluminada con luz artificial y sin luz natural [se descubrirían sólo los casos más extremos y no los leves] o medir la tensión arterial, justo después de que el paciente subió las escaleras [siempre arrojaría valores más altos]), con problemas de calibración de los instrumentos o cuando los evaluadores usan diferentes criterios técnicos para la medición. El sesgo invalida las conclusiones de un estudio (afecta la validez interna), su efecto suele ser irreducible y no siempre es posible medirlo. El sesgo puede evitarse con un diseño apropiado del estudio de investigación y con atención a los detalles técnicos de la medición.

## Tipos de sesgo

**Malacласificación no diferencial y diferencial.** Cuando la magnitud del sesgo en la medición de la variable de exposición no se asocia con el valor de la variable de resultado (es decir, la diferencia se mantiene constante), se dice que ocurrió malaclasificación no diferencial, mientras que cuando el tamaño del sesgo se asocia con la magnitud de la variable de exposición, se dice que ocurrió malaclasificación diferencial.

La malaclasificación no diferencial siempre tiene el mismo efecto; reduce la fuerza de la asociación, es decir, disminuye la razón de momios o el riesgo relativo (tal como lo haría el error aleatorio), mientras que la malaclasificación diferencial puede tener cualquier efecto en la razón de momios (puede reducirla, aumentarla o cambiar de dirección, causando asociaciones espurias).

Un ejemplo de malaclasificación diferencial ocurrió en estudios de casos y controles de malformaciones congénitas, donde la posibilidad de recordar la exposición a potenciales sustancias nocivas dependía de que la madre fuera caso (lo recuerda más) o control (tenía menos interés en buscar ese dato en su memoria). Si todas las pacientes tuvieran la misma baja probabilidad de recordar esa exposición, independientemente de que sean casos o controles, aún sería un sesgo, pero la malaclasificación sería no diferencial.

**Sesgo de selección.** Se define como una distorsión en el efecto observado (la razón de momios o el riesgo relativo), ocasionada por los procedimientos para reclutar o seleccionar a los sujetos del estudio. Se han mencionado los siguientes tipos (el listado no es exhaustivo):

- a) *Filtro de referencia y popularidad*. Algunos centros seleccionan el tipo de pacientes que atienden y las características de estos enfermos podrían ser distintas del enfermo típico.
- b) *Incidencia-prevalencia* (Neyman). El estudio de casos prevalentes puede distorsionar la imagen de la enfermedad; cuando ésta tiene una alta letalidad en las etapas tempranas, los casos disponibles (los prevalentes) no incluyen aquellos que murieron en las etapas iniciales de la enfermedad y esto puede causar ¡que los factores de riesgo aparezcan como protectores!
- c) *No respondedor, membresía, voluntario, trabajador sano*. Algunas personas pueden “autoseleccionarse” para un estudio. Por ejemplo, cuando la exposición tiene una connotación moral, los expuestos suelen negarse a participar; algunas personas con una mayor preocupación por su salud (y diferente “estilo de vida”) suelen estar más dispuestas a participar en un estudio; algunos oficios y profesiones atraen personas con características distintas (mejor salud) que los de la población general.
- d) *Susceptibilidad*. Algunas personas se saben con mayor riesgo de enfermedad por su estilo de vida y están más dispuestas a participar en un estudio en el que se les hará la prueba diagnóstica que define si ya están enfermos. Algunas ocupaciones suponen mayor riesgo de ciertas enfermedades; si uno estudió sólo a esas personas, las tasas de incidencia pueden ser mayores que las de la población.
- e) *Berkson*. Ocurre en estudios de casos y controles; se descubrió cuando se observó asociación entre dos enfermedades sólo en pacientes hospitalizados y no en las personas que viven en la comunidad. La razón de este sesgo fue que cuando estas dos enfermedades se encontraban en un paciente, era más probable que se le hospitalizara, que cuando estaba presente cualquiera de ellas por separado. La supuesta asociación no era real y estaba causada por los criterios de admisión del hospital. Así, los criterios de admisión de un hospital pueden causar asociaciones espurias.

**Sesgo de información.** Se refiere al sesgo de medición, cuando los resultados del proceso arrojan valores que son sistemáticamente diferentes del valor real. Se refieren los siguientes:

- a) *Expectativa diagnóstica*. Ocurre cuando quien obtiene los datos anota el valor que quisiera o esperaba observar y no el que realmente ocurrió (revise las últimas 10 mediciones de tensión arterial que estén a su disposición; ¡ahora trate de explicar por qué todas acaban en 0 o 5 [135/70] y no en cualquier otro número [132/73]!).
- b) *Inferencia*. Cuando se registra lo que se cree que ocurrió y no lo que se vio. El enfermo informa orina roja y el médico anota “hematuria” (este enfermo podría estar tomando antituberculosos y no tendría “hematuria”, a pesar del color de su orina).
- c) *Sospecha diagnóstica o de exposición*. Un médico busca con más acuciosidad un posible efecto, cuando sabe que el paciente está expuesto a sus factores de riesgo; en consecuencia, se descubre más a menudo ese resultado en los expuestos que en los no expuestos (en quienes no se buscó con el mismo interés). Aunque esa es la práctica clínica recomendable en la consulta habitual, en los estudios de investigación las

mediciones deben hacerse con la misma frecuencia y acuciosidad en los expuestos y en los no expuestos.

- d) *Recuerdo, rumiación.* Un problema frecuente en los estudios de casos y controles. Es más probable que quien tiene la enfermedad hará el intento de recordar su exposición a los factores de riesgo que quien no la tiene y, en consecuencia, se detectan más casos expuestos que controles (que también podrían estar expuestos, pero no lo recuerdan).
- e) *El enfermo “obsequioso”.* Los enfermos muestran su “interés” por participar activamente en su curación y premian los esfuerzos de su médico, informando molestias de menor intensidad de las que en verdad tienen.
- f) *Deseabilidad social.* Las personas no desean ser vistas como alguien que hace “cosas” socialmente mal vistas y, en consecuencia, responden de manera sesgada. Por ejemplo, las preguntas relacionadas con la cantidad de alimentos que se consumen en un día, la ingesta de alcohol y otros psicotrópicos, las preferencias sexuales y la violencia familiar o de género, suelen generar respuestas que no corresponden con la cantidad real que toma la persona con sus preferencias o su verdadera conducta.
- g) *Prevaricación.* Podría corresponder más a fraude que a sesgo, si el enfermo espera una “ganancia” (compensación, certificado de incapacidad, etc.) si informa molestias más intensas.
- h) *Entrevistador.* Cuando por la manera de hacer las preguntas se induce una forma de respuesta que no corresponde con lo que el paciente contestaría si la pregunta se hubiera hecho de manera neutra.

**Sesgo de confusión.** Se dice que ocurrió sesgo de confusión cuando se mezclan los efectos de la variable de exposición con una variable extraña y con los del resultado. Por ejemplo, un grupo de investigación informó que el consumo de café se asociaba con la presencia de una forma de cáncer. Otros investigadores no pudieron replicar ese resultado. Cuando se investigaron las causas de esta discrepancia se observó que entre los tomadores de café también había más fumadores y que la mayor incidencia de cáncer ocurría en los fumadores. Así, la supuesta asociación entre consumo de café y cáncer ocurrió por un sesgo de confusión; es decir, el consumo de café se asociaba con la exposición (fumar) y con el efecto (cáncer), pero no era parte del mecanismo de acción. Causa una asociación que no existe. Para que una variable se considere confusora se requiere que ésta se asocie con la enfermedad aun en ausencia de la exposición y que se asocie con la exposición; pero que no sea consecuencia de la exposición.

## Control del sesgo

A diferencia del error aleatorio, el sesgo difícilmente puede medirse y cuando ocurre, ya no existe forma de atenuar sus consecuencias. Sólo puede evitarse el sesgo con un apropiado diseño del proyecto de investigación. Para evitar el sesgo de selección debe ponerse especial atención a los criterios de inclusión y a los procedimientos de reclutamiento de pacientes, con la finalidad de obtener una muestra de enfermos que sean lo más parecidos posibles al paciente típico con ese trastorno o enfermedad.

A fin de evitar el sesgo de información debe ponerse especial cuidado en los procedimientos de medición; por ejemplo, aplicar el mismo procedimiento a todos los sujetos, anotar lo observado y las inferencias y distinguirlos (orina roja, probablemente hematuria), hacer mediciones de manera independiente y a ciegas (para evitar que el conocimiento de las hipótesis o de la exposición induzcan al examinador a buscar con más ahínco un resultado en ciertos pacientes y no en otros). Es primordial consignar preguntas objetivas y sin juicios de valor (p. ej., indagar cuántas bebidas alcohólicas consumió el paciente en la última semana, en vez de preguntarle si “bebe en exceso”). Es fundamental diseñar definiciones operativas para el procedimiento de medición y para la interpretación de los resultados, que sean explícitas y no dependan de la interpretación personal; asimismo, se debe entrenar y estandarizar a los observadores de manera que obtengan resultados iguales en pacientes iguales, etcétera.

A fin de evitar el sesgo de confusión se han usado diferentes procedimientos: restricción (p. ej., si la edad es un confusor, es conveniente limitar el tipo de pacientes a sólo ciertas edades, y evitar reclutar a los muy jóvenes o los muy viejos); “apareamiento” (p. ej., en un estudio de casos y controles, se debe elegir un control hombre por cada caso hombre); estratificación (p. ej., elegir proporciones iguales de fumadores y no fumadores en los expuestos, y en los no expuestos o en el análisis estadístico, aplicar un procedimiento que compare sólo los fumadores entre expuestos y no expuestos y luego haga lo mismo con los no fumadores); ajuste de tasas (ajustar las tasas con respecto de una población típica —ajuste directo— o usar incidencia esperada con ajuste indirecto); ajuste multivariado, para considerar simultáneamente los efectos de múltiples variables y, por último, si el diseño es de un ensayo clínico controlado, se deben asignar los tratamientos al azar.

## Evaluación de la concordancia

El atributo más importante de una medición es la consistencia. Imagine qué clase de ciencia se haría si cada vez que una nueva persona revisa la misma evidencia llega a conclusiones diferentes. La concordancia sólo puede evaluarse si se hacen mediciones por duplicado, sea que un mismo observador haga la medición en dos ocasiones diferentes o que dos observadores hagan la medición del mismo sujeto o muestra. Cuando se trata de muestras biológicas o biopsias, esto no es tan difícil, pues se pueden almacenar para reanalizarlas después; pero cuando se trata de fenómenos clínicos es importante considerar que el periodo entre la primera y la segunda mediciones sea suficientemente largo como para que el individuo no recuerde lo que respondió la primera ocasión, mas no tan largo como para que sus condiciones clínicas (y el resultado esperado de la medición) hubieran cambiado.

Una vez que tiene mediciones por duplicado, debe elegir cómo analizarlas. Algunos investigadores comparan los resultados de la primera y la segunda mediciones con pruebas de hipótesis (p. ej., comparan promedios con la prueba  $t$  de Student o proporciones con la prueba  $\chi^2$ ). Esta estrategia es inapropiada. Si la prueba no muestra diferencias “estadísticamente significativas” (p. ej.,  $p > 0.05$ ), esto no significa que los valores de la primera y la segunda mediciones concuerden. Una diferencia pequeña, pero clínicamen-

CV = D.E./promedio

| Muestras de un mismo paciente | Reflectómetro | Laboratorio     |
|-------------------------------|---------------|-----------------|
| 1                             | 140           | 134             |
| 2                             | 157           | 129             |
| 3                             | 120           | 132             |
| 4                             | 115           | 133             |
| 5                             | 160           | 134             |
| Promedio D.E.                 | 138.4         | 132.4           |
|                               | 20.6          | 2.07            |
| Coefficiente de variación     | 0.14<br>(14%) | 0.015<br>(1.5%) |

Figura 9-10. Coeficiente de variación.

te importante, puede no ser detectable con una estrategia así. Otra estrategia inadecuada sería calcular la “correlación” entre los dos resultados (p. ej., con el coeficiente  $r$  de Pearson). El problema es que las pruebas de correlación están diseñadas para una “tendencia”; es decir, que la relación entre dos mediciones se puede explicar con la ecuación de la línea recta. Si la segunda medición muestra valores consistentemente más altos o más bajos que la primera, el coeficiente  $r$ , arrojará valores altos, en tanto la relación entre esas dos mediciones siga siendo “lineal”. El uso del coeficiente  $r^2$  no resuelve el problema, pues éste señala en qué proporción se explica una unidad de cambio en la primera medición, con respecto de la segunda. Para conocer la concordancia es indispensable aplicar pruebas específicas que se muestran a continuación.

**Coeficiente de variación.** Este coeficiente arroja como resultado una medida relativa de la dispersión de los datos alrededor del promedio; para obtenerlo es necesario hacer varias mediciones del mismo sujeto o muestra. Así, se calculan el promedio y desviación estándar. Después se divide la desviación estándar entre el promedio y el resultado se multiplica por una base (p. ej., 100, para expresarlo como porcentaje) (figura 9-10).

No existe una regla fija para interpretar el coeficiente de variación pero, en general, se estima que un coeficiente mayor al 10% suele mostrar una variación excesiva.

**Prueba kappa ( $\kappa$ ).** Suponga que dos médicos examinan a los mismos 100 pacientes y sus diagnósticos concuerdan en 75 de ellos. En el resto, el diagnóstico de un médico es diferente del otro. ¿Cómo juzgaría este grado de “acuerdo” o concordancia? Usted puede examinar este grado de acuerdo general (75%) o buscar el acuerdo específico (qué porcentaje tiene la enfermedad); pero aun esto no resuelve su problema. Para juzgar el grado

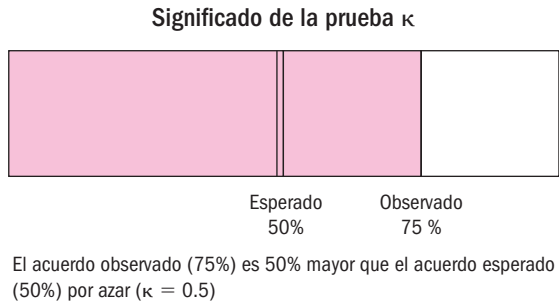


**Figura 9-11.** Interpretación del coeficiente kappa.

de concordancia, querría saber si estos médicos se ponen de acuerdo sólo por efecto del azar o si concuerdan en sus diagnósticos, más allá de lo esperado por efecto del azar. El coeficiente kappa se calcula con base en la siguiente fórmula:

$$\kappa = \frac{\text{Concordancia observada} - \text{concordancia esperada}}{1 - \text{concordancia observada}}$$

El numerador representa qué tanto excede la concordancia observada de la esperada por azar y el denominador representa qué tanto desacuerdo era esperable por efecto del azar. El coeficiente kappa siempre arroja valores entre 1.0 y  $-1.0$ . Un resultado de 0 (cero) no implica que no hubo acuerdo; más bien, señala que el acuerdo observado no excede al esperado por azar. Con este coeficiente se juzga en qué proporción el acuerdo observado excede al esperado por efecto del azar (los valores negativos se refieren a desacuerdo); mientras más cercano a 1.0 sea el resultado, más excede el acuerdo observado al esperado y más consistente se considera la medición (figura 9-11). Landis y Koch propusieron el esquema que se muestra en la figura 9-12 para juzgar el resultado de este coeficiente.

| Interpretación de la prueba $\kappa$ |           |
|--------------------------------------|-----------|
| 0.00 a 0.20                          | Mala      |
| 0.21 a 0.40                          | Regular   |
| 0.41 a 0.60                          | Buena     |
| 0.61 a 0.80                          | Muy buena |
| 0.81 a 0.99                          | Excelente |
| 1.00                                 | Perfecta  |

**Figura 9- 12.** Clasificación de Landis y Koch para el coeficiente kappa.

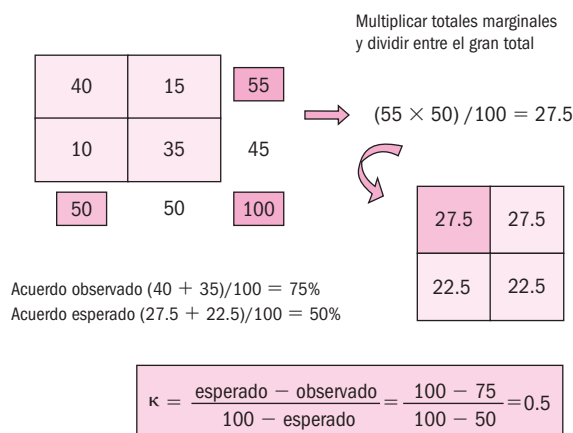
Cálculo valores esperados prueba  $\kappa$ 

Figura 9-13. Cálculo del coeficiente kappa.

Este coeficiente se usa sólo en los casos en los que el resultado de la medición se expresa en una variable nominal (p. ej., cuando los resultados se expresan como presente, ausente), y para estimar las frecuencias esperadas por azar se puede usar el mismo procedimiento que se usaría en una prueba  $\chi^2$ . Es decir, para calcular la frecuencia esperada en una casilla, se multiplican los totales de la suma de la columna correspondiente a esa casilla con el de la suma del renglón al que corresponde esa casilla y este producto se divide entre el total de la suma de las cuatro casillas. Una vez calculada la frecuencia esperada en cada casilla, se suman las correspondientes al acuerdo (aquellas que corresponden a ++ y --) y se dividen entre el total de las cuatro casillas; esta es la proporción de acuerdo esperada (figura 9-13).

La prueba kappa está asociada a la distribución  $\chi^2$ . Así que cuando esta distribución no es aplicable (p. ej., cuando la frecuencia esperada de alguna de las casillas es menor de 5) se usa la  $\chi^2$ . En este ejemplo, el resultado del coeficiente kappa fue  $\kappa = 0.5$ ; éste significa que el acuerdo observado de 75% excede en 50% al acuerdo esperado por azar y se le considera “bueno”.

**Prueba kappa ponderada ( $\kappa_w$ ).** Se usa cuando el resultado de la medición se expresa en una variable ordinal (p. ej., 0, +, ++). En estos casos, el desacuerdo puede ser pequeño (p. ej., entre 0 y +) o mayor (p. ej., entre 0 y ++) y en consecuencia se considera que un desacuerdo más grande debe tener un valor o peso (“weight” o “W” en inglés) mayor en el resultado final del coeficiente. Aunque el número de casillas es mayor que al calcular la prueba  $\kappa$  (en este ejemplo sería  $3 \times 3$ ), las frecuencias esperadas se calculan de la misma manera (totales marginales —es decir columna por renglón— entre el gran total). Luego se asigna “peso” al desacuerdo (usualmente se estima como número de casillas,  $\kappa - 1$ , en este caso  $w + 3 - 1 = 2$ ). En los casos donde los observadores concuerdan, el peso sería  $w = 0$ , en donde el desacuerdo es sólo entre 0 y +, el peso  $w = 1$  al final, donde

**Kappa ponderada ( $\kappa_w$ )**

| Valores observados |         |      |       |
|--------------------|---------|------|-------|
|                    | Ausente | Leve | Grave |
| Ausente            | 11      | 9    | 1     |
| Leve               | 6       | 6    | 7     |
| Grave              | 3       | 7    | 11    |

| Valores esperados |         |      |       |
|-------------------|---------|------|-------|
|                   | Ausente | Leve | Grave |
| Ausente           | 6.9     | 7.6  | 6.5   |
| Leve              | 6.2     | 6.9  | 5.9   |
| Grave             | 6.9     | 7.6  | 6.5   |

Peso ( $w = \kappa - 1$ ); en este caso  $\kappa = 3$ ;  $w = 2$

| Peso ( $w$ ) para cada casilla |         |      |       |
|--------------------------------|---------|------|-------|
|                                | Ausente | Leve | Grave |
| Ausente                        | 0       | 1    | 2     |
| Leve                           | 1       | 0    | 1     |
| Grave                          | 2       | 1    | 0     |

$$\kappa_w = (1 - (\sum f_{0w}) / (\sum f_{ew}))$$

$$\kappa_w = 1 - (((9+6+7+7)*1)+((1+3)*2)) / (((7.6+6.2+5.9+7.6)*1)+((6.9+6.5)*2))$$

$$\kappa_w = 1 - (0.68) = 0.31$$

**Figura 9-14.** La prueba kappa ponderada ( $\kappa_w$ ).

el desacuerdo es mayor (entre 0 y ++), el peso sería  $w = 2$ . La fórmula para calcular el coeficiente  $\kappa_w$  es la siguiente:

$$\kappa_w = (1 - (\sum f_{0w}) / (\sum f_{ew}))$$

Donde  $\sum f_{0w}$  es la suma de las frecuencias observadas, multiplicadas por su peso (0, 1 o 2) y  $\sum f_{ew}$  es la suma de las frecuencias esperadas, multiplicadas por su peso.

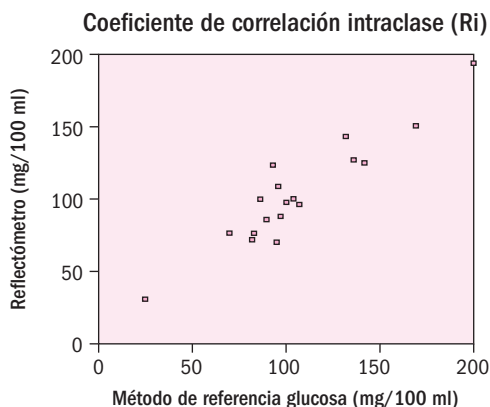
Note que aquí no se calculan las proporciones de acuerdo y se usan directamente las frecuencias observadas y esperadas y que como aquellas casillas donde hubo acuerdo se multiplican por 0 (cero), sólo aparecen en la fórmula las casillas donde hubo desacuerdo. Por eso, el resultado de la división debe restarse de 1.0, para expresarlo nuevamente como un coeficiente de acuerdo (figura 9-14).

El resultado de  $\kappa_w$  suele juzgarse más estrictamente que la prueba  $\kappa$ , porque ya se dio un peso más grande a un desacuerdo mayor y suele ser necesario que  $\kappa_w \geq 0.7$  para considerar que la concordancia es “buena”.

**Coefficiente de correlación intraclase (Ri).** Cuando la variable es continua, sería imposible aplicar la prueba  $\kappa_w$  porque habría que construir una tabla con demasiadas casillas y algunas de ellas quedarían vacías. En un caso así, se aprovechan las pruebas de análisis de la varianza. Explicar una prueba así estaría fuera del alcance de este capítulo, pero baste decir que en una prueba de esta naturaleza pueden estimarse las distintas fuentes de variación: la varianza total, la que se debe a los sujetos, la relacionada con las repeticiones de la medición (método) y la varianza del error. La fórmula general para el coeficiente de correlación intraclase es la siguiente:

$$Ri = (\sigma^2 \text{ individuos} - \sigma^2 \text{ error}) / (\sigma^2 \text{ individuos} + \sigma^2 \text{ error} + 2\sigma^2 \text{ método})$$





| Fuente de variación                                                                                                                                         | Suma de cuadrados | Grados de libertad | Cuadrado medio | "F"            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------------|----------------|----------------|
| Total                                                                                                                                                       | 396039.4          | 1                  | 396039.4       |                |
| Sujetos                                                                                                                                                     | 47072.1           | 17                 | 2768.9         | 29.23          |
| Métodos                                                                                                                                                     | 40.3              | 1                  | 40.3           | 0.42           |
| Error                                                                                                                                                       | 1610.08           | 17                 | 94.7           | ( $p = 0.52$ ) |
| $Ri = (CMs - CMe) / (CMs + CMe + (2(CMm)))$ $Ri = (2\,768.9 - 94.7) / (2\,768.9 + 94.7 + (2 + (40.3)))$ $Ri = 0.96 \text{ (IC}_{95\%} \text{ 0.90 - 0.98)}$ |                   |                    |                |                |

**Figura 9-15.** Coeficiente de correlación intraclase (Ri).

Donde  $\sigma^2$  significa varianza y se identifican sus fuentes (individuos, las diferencias de un sujeto a otro; método, las diferencias entre la primera y la segunda mediciones de cada individuo; error, una parte de la variabilidad en la que no podemos identificar su causa). De hecho, si se realiza una prueba  $\kappa_w$  con estos mismos datos, pero elevando al cuadrado el peso (w), obtendría el mismo resultado que con Ri. Este coeficiente arroja resultados entre  $-1.0$  y  $+1.0$ , donde 0 (cero) significa que la concordancia no es superior a la esperada por efecto del azar (figura 9-15).

**Método de Bland y Altman.** Todos los procedimientos anteriores se expresan como coeficientes, que por definición no tienen unidades de medición. Son útiles para comparar diferentes métodos que usan unidades de medición diferentes y para economizar espacio (en lugar de describir todos los pormenores, podría resumirse diciendo que la concordancia fue  $\kappa = 0.82$ ). Con frecuencia no basta saber qué tan repetible es una medición. También es preciso determinar cuánto difiere una medición de otra.

Considere, por ejemplo, una báscula para medir el peso de un adulto. En esa báscula el interés reside en conocer diferencias de al menos 500 g entre una y otra mediciones. Diferencias más pequeñas suelen ser irrelevantes si la persona originalmente pesa 80 kg. Pero si la báscula se usa para pesar recién nacidos, quizá sea importante detectar cambios tan

| Paciente | Laboratorio | Reflectómetro | Diferencia |
|----------|-------------|---------------|------------|
| 1        | 104         | 99,9          | 4,1        |
| 2        | 82          | 72            | 10         |
| 3        | 90          | 86            | 4          |
| 4        | 83          | 77            | 6          |
| 5        | 136         | 127           | 9          |
| 6        | 25          | 31            | -6         |
| 7        | 86          | 100           | -14        |
| 8        | 100         | 98            | 2          |
| 9        | 169         | 150           | 19         |
| 10       | 95          | 70            | -35        |
| 11       | 142         | 125           | 17         |
| 12       | 96          | 109           | -13        |
| 13       | 93          | 124           | -41        |
| 14       | 200         | 194           | 6          |
| 15       | 97          | 88            | 9          |
| 16       | 70          | 77            | -7         |
| 17       | 132         | 144           | -12        |
| 18       | 107         | 97            | 10         |
| Prom.    | 105.9       | 103.8         | -2.1       |
| D.E.     | 38.9        | 36.6          | 26.8       |

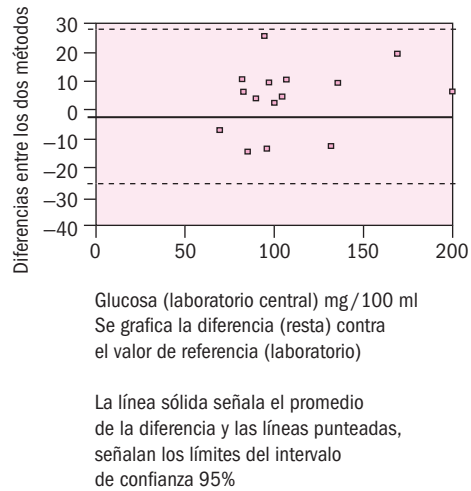


**Figura 9-16.** Método de Bland y Altman.

pequeños como 30 g y, aunque el coeficiente de correlación intraclass parezca apropiado (p. ej., 0.9), la báscula sería inapropiada si en promedio la diferencia entre la primera y la segunda mediciones excede 30 g. Para evaluar la repetibilidad de manera cuantitativa, Bland y Altman propusieron, en 1983, restar las dos mediciones y graficar el valor de una de ellas (X) con el valor de la resta (Y). Con un gráfico así, se pueden evaluar la diferencia promedio entre las dos mediciones, la dispersión de las diferencias y el límite de detección (promedio de la resta  $\pm 1.96$  veces su desviación estándar) y si es que existe sesgo en función del valor original de X (figura 9-16).

## Homogeneidad y otros procedimientos para evaluar confiabilidad

**Prueba  $\alpha$  de Cronbach.** Muchas de las cosas que se miden en Medicina no se pueden preguntar de manera explícita (p. ej., ¿cuál es su calidad de vida?). Es preciso hacer cuestionarios con diferentes preguntas (p. ej., dolor, emociones, funcionalidad, etc.). Cada una de ellas arroja su propia respuesta y el resultado final suele ser la suma de todas ellas (como sucedió en el ejemplo del APGAR). Para usar esa suma como una representación de la “cantidad” de un atributo concreto, todas estas preguntas deben tener consistencia interna u homogeneidad; es decir, todas deben medir diferentes aspectos de lo mismo. Deben relacionarse una con otra y con la suma final.

La fórmula para calcular el coeficiente alfa ( $\alpha$ ) es la siguiente:

$$\alpha = (\kappa/\kappa - 1)(1 - (\sum \sigma^2_i/\sigma^2_T))$$

Donde  $\kappa$  es el número de preguntas del cuestionario o ítems,  $\sum \sigma^2_i$  es la suma de la varianza de cada uno de los reactivos y  $\sigma^2_T$  es la varianza de la suma de todas las pre-

guntas. Si la suma de las varianzas de cada pregunta y la varianza del puntaje total son iguales, el resultado de dividir las será 1.0 e indicará la más alta homogeneidad posible. Si el resultado es 0 (cero), eso indica que las preguntas no están relacionadas unas con otras y que no puede usarse la suma de cada una de las respuestas, aunque podrían analizarse por separado las respuestas de cada pregunta. La prueba  $\alpha$  está diseñada para evaluar respuestas que se expresan de manera cuantitativa (como la suma de puntos; por ejemplo con valores entre 0 y 10). Para evaluar cuestionarios que se responden como sí (1) o no (0), podría usarse la prueba 20-20 de Kuder-Richardson, que es muy similar. El coeficiente que resulta de aplicar la prueba  $\alpha$  es idéntico a lo que resultaría de estimar el promedio de todas las posibles correlaciones entre los ítems y su total. Cuando el cuestionario se usa para analizar grupos, basta un coeficiente  $\alpha \geq 0.7$ . Cuando se usan para trabajo clínico con pacientes individuales, se espera que el coeficiente  $\alpha \geq 0.9$ .

Note cómo esta fórmula corresponde con la definición de confiabilidad que se hizo en una sección previa (la división de una fuente de variación [en este caso los reactivos]) entre la varianza total y cómo este coeficiente se parece al coeficiente de correlación intraclass (que, de hecho, también tiene que ver con la prueba  $\kappa$ ). Lo que cambia es el numerador (los resultados en dos mediciones o los resultados de cada pregunta). Lo que hace a estas pruebas ( $\kappa$ ,  $\alpha$ , etc.) muy populares es que pueden hacerse con un mínimo de conocimientos de matemáticas y hasta con la calculadora de su teléfono celular; además, están en casi todos los paquetes de software de análisis estadístico. En los últimos años se han diseñado otras pruebas para evaluar aspectos semejantes (teoría de la generalizabilidad [G], modelamiento multidimensional, estructura multidimensional latente, análisis factorial, entre otras). Todas ellas se derivaron del análisis de la varianza, requieren de equipo y software más complejo y no se revisarán en este capítulo.

## Teoría de la “sensibilidad”

Con frecuencia se requiere una escala o cuestionario para hacer una medición específica. Para resolver esta necesidad el investigador revisa lo que hay publicado al respecto y no es raro encontrar más de un cuestionario que pudiera ser útil. Feinstein propuso el siguiente sistema para evaluar cuál de ellas sería la mejor para cubrir sus necesidades: lo llamó “teoría de la ‘sensibilidad’”, que se refiere más al hecho de que la medición tenga sentido (*sense* en inglés) que al concepto de sensibilidad de las pruebas diagnósticas. De hecho, es recomendable aplicar esta evaluación a cada instrumento de medición que consideremos aplicar en nuestros pacientes, antes de usarlo. Note que esta evaluación no requiere de usar ninguna prueba estadística; más bien se trata de un juicio personal.

- a) La escala debe ser apropiada al propósito de la medición y a su marco de referencia: si se desea evaluar el pronóstico, no debe aplicarse una escala diseñada como instrumento para clasificación. Además, si se diseñó para los pacientes de la terapia intensiva, no necesariamente funcionará bien en los pacientes de la consulta externa (seguramente no podrá detectar diferencias pequeñas).

## Ejemplos de escalas para responder preguntas en cuestionarios

Continua: 145 mmHg

Categoría:                
                    Sí      No

Tipo Likert:

☒ Desacuerdo  
☐ Desacuerdo parcial  
☐ No sé  
☐ Acuerdo parcial  
☐ De acuerdo

### Escala visual análoga

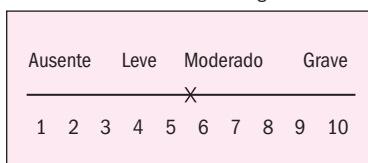


**Figura 9-17.** Se muestran cuatro formatos de respuesta: de respuesta abierta para variables continuas, de respuesta cerrada para variables categóricas, un formato de tipo Likert y otro de una escala visual análoga.

- b) Debe contener todas las preguntas o reactivos que sean importantes y cada pregunta que contenga la escala debe justificarse clínicamente con base en la fisiopatología, connotación biológica, social, etcétera.
- c) Debe tener la posibilidad de ofrecer resultados consistentes, es decir, debe haber instrucciones de cómo aplicarla y evaluarla y, además, las preguntas y las opciones de respuesta deberán estar redactadas de manera tal, que permitan que se respondan sin sesgo (p. ej., sin dobles negativas, doble pregunta, etc.).
- d) Las escalas de medición de cada pregunta deben permitir que se distinga un fenómeno de interés clínico o una diferencia importante desde el punto de vista clínico (figura 9-17).
- e) Debe tener validez de apariencia, es decir, debe evaluar la evidencia apropiada; sus componentes deben ser coherentes, debe considerarse, si es el caso, la participación del paciente; por ejemplo, si se requiere que el paciente haga alguna tarea, debe asegurarse que desempeñó esa tarea con el esfuerzo requerido (la prueba de esfuerzo requiere que el paciente alcance al menos 80% de la frecuencia cardíaca esperada para su edad para que sea interpretable).
- f) Debe tener validez de contenido; es decir, los datos que arroje deben ser de alta calidad, deben estar representados todos los componentes del concepto o idea que pretende medir y cada pregunta deberá tener un peso adecuado en el puntaje final, en proporción con su importancia relativa.
- g) Debe ser de fácil aplicación.

## Adaptación y traducción de cuestionarios y escalas para uso local

Si después de aplicar su análisis de sensibilidad el instrumento muestra fallas, es factible modificarlo para que funcione como lo requiere (p. ej., podría cambiarse la escala de me-

dición de una o más preguntas, agregar un manual con instrucciones específicas de cómo hacer las mediciones, etc.). De cualquier manera, si el investigador modificó este cuestionario, debe ponerlo a prueba y demostrar que esa modificación funciona en forma adecuada.

Si lo que usted encontró está escrito en otro idioma, debe traducirlo al idioma en el que habrá de usarse; para que esta traducción funcione de manera correcta, es preciso seguir los siguientes pasos:

- a) Traducir del idioma original al idioma “blanco” (en el que habrá de usarse). Es conveniente que el traductor de esta porción, tenga como idioma nativo, el idioma “blanco” (para este ejemplo, que su idioma materno sea español).
- b) Retraducir la versión traducida al idioma original. Esta parte deberá hacerla un traductor que tenga como idioma nativo, el idioma original.
- c) Comparar las dos versiones (la original y la retraducida) en el idioma original y decidir si son equivalentes.
- d) Evaluar el “contenido cognitivo” (el significado de cada pregunta y de las opciones de respuesta), comparando la versión original con la versión traducida al idioma blanco. Esto es de particular importancia porque la traducción no puede ser literal. Debe ser gramaticalmente correcta, pero las preguntas y las opciones de respuesta deben tener el mismo significado que en el idioma original (imagine cómo traduciría, en un cuestionario de depresión, “*I am feeling blue*”).
- e) Por último, debe verificarse que las características de las escala se mantienen al aplicarla en el nuevo idioma (estructura factorial,  $\alpha$ , etc.).

Sólo entonces será posible aplicar este cuestionario, con la confianza de que sus resultados serán equivalentes a los del cuestionario en su idioma original.

## Construcción de índices o escalas de medición

Siempre será más práctico usar un cuestionario o escala existente, que desarrollar uno nuevo; sin embargo, existe la posibilidad de que usted necesite medir algo para lo que aún no existe un instrumento apropiado. En un caso así, tendría que desarrollarlo; en estos casos, se recomienda seguir los siguientes pasos:

1. Defina conceptualmente aquello que quiere medir.
2. Identifique qué dominios lo conforman.
3. Construya un listado de reactivos o preguntas para cada uno de los dominios; en esta etapa, es conveniente que haga preguntas de más.
4. Asigne una escala de medición a cada reactivo o pregunta.
5. Decida cómo se agrupan los valores de resultado de cada reactivo o pregunta, en una calificación final (figura 9-18).
6. Proceda a reducir el número de reactivos o preguntas redundantes en cada dominio (para esto puede usar la prueba  $\alpha$  de Cronbach).
7. Evalúe sensibilidad, consistencia y validez.

### Cómo se organizan los puntajes de un cuestionario

- **Suma algebraica:**
  - Total = respuesta 1 + Respuesta 2 +...
  - La mayoría de las escalas usa este procedimiento, como la escala Apgar
  - La suma puede ser ponderada: (respuesta 1\*0.5) + (respuestas 2\*0.7)...
- **Perfil:**
  - Subescala 1, subescala 2, subescala 3...
  - Un ejemplo es el inventario multifásico de la personalidad de Minnesota, que informa cada subescala (esquizoide, paranoide, etc. ) por separado
- **Escala jerárquica:**
  - Un ejemplo es la clasificación T,N,M; del cáncer, que se convierte en estadios  
Estadio 0 = T1,N0,M0; Estadio 2 = T1,N1,M0,... Estadio 4 = cualquier T, cualquier N, M<sub>1</sub>
- **Conglomerados booleanos:**
  - Se informa qué combinaciones de características tiene cada sujeto (A, B, C, AB, AC, BC, ABC), como en la teoría de conjuntos

**Figura 9-18.** Cómo organizar el puntaje final de un cuestionario.

A fin de llevar a cabo este último procedimiento es necesario poner a prueba el cuestionario o escala, en pruebas de campo; es decir, hay que aplicarlo en algún número de personas para ver cómo funciona y refinar su funcionamiento. Esto debe hacerse tantas veces como sea necesario, para que el instrumento opere con las características deseadas. No todos los pasos deberán repetirse cada vez. Pero, si no funciona como se espera, quizá sea necesario reescribir alguna o más preguntas u opciones de respuesta y probar cómo funcionan estos cambios.

### Algunas recomendaciones adicionales para obtener información certera de sus pacientes durante el examen físico

- a) Asegure que el medio es apropiado para hacer la medición (para detectar ictericia es necesario contar con luz natural; una buena auscultación requiere que no haya ruido ambiental).
- b) Corrobore sus hallazgos clave:
  - Repita la medición hasta estar seguro.
  - Corrobore sus hallazgos con los registros (expediente, notas) y testigos (¿cuándo apareció esto?).
  - Corrobore el hallazgo con las pruebas apropiadas (ictericia = medición de bilirrubinas).
  - Solicite corroboración independiente con un colega (“dime cómo lo valoras”, y no “dime si lo ves ictérico”).
- c) Registre evidencia e inferencias y distíngalas con claridad (recuerde que hay muchas cosas que pueden cambiar el color de la orina, no sólo la hematuria).

- d) Utilice los auxiliares apropiados (todo médico necesita su estetoscopio y su estuche de diagnóstico).
- e) Interprete los exámenes auxiliares de manera independiente (por ejemplo, primero describa para usted qué vio en la radiografía y sólo después, trate de correlacionar sus hallazgos con la sintomatología y los resultados del examen físico).
- f) Aplique ciencias sociales (recuerde que no todas las personas responden igual y que la cultura matiza la forma en que los pacientes expresan sus molestias: “me duele” también puede ser “me ofende”, “me recuerda”, “me escuece” y así por el estilo).

---

## Bibliografía

- Armstrong B, White E, Saracci R. *Principles of exposure measurement in epidemiology*. Monographs in Epidemiology and Biostatistics, No. 21. Oxford University Press, Oxford, 1994.
- DeVellis R. Scale development: Theory and applications. Applied Social Research Methods. Series No. 26, SAGE publications, Inc. Newbury Park, Cal. 1991.
- DeVon H, Block M, Moyle-Right P, Ernst D, Haydn S, Lazzara D, Savoy S, Kostas-Polson EA. Psychometric toolbox for testing validity and reliability. *J Of Nurs Scholarship* 39:2, 155-164. 2007.
- Dunn G. Design and analysis of reliability studies. Oxford University Press, Nueva York, EUA. 1989.
- Elliot D, Hickam D. Evaluation of physical examination skills reliability of faculty observers and patient instructors. *JAMA* 258:3405-3408. 1987.
- Feinstein A. *Clinimetrics*. Yale University Press, New Haven, CT, EUA. 1987.
- Gustafson P. *Measurement error and misclassification in statistics and epidemiology. Impacts and bayesian adjustments*. Chapman & Hall/CRC, Boca Raton, EUA. 2004.
- Hulley S, Cummings S, Browner W, Grady D, Hearst N, Newman T. Designing clinical research, 2a. ed. Lippincott Williams & Wilkins, Filadelfia, EUA. 2001.
- Kelsey J, Whitmore A, Evans A, Douglas-Thompson W. *Methods in observational epidemiology*, 2a. ed. Oxford University Press, Nueva York, EUA. 1996.
- Nunnally J, Bernstein I. Psychometric theory, 3a, ed. *McGraw-Hill Series in Psychology*, McGraw-Hill, Nueva York, EUA. 1994.
- Schlesselman J. *Case control studies: design, conduct, analysis*. Oxford University Press, Nueva York, EUA. 1982.
- Streiner D, Norman G. *Health measurement scales: A practical guide to their development and use*. 2a. ed. Oxford University Press, Nueva York, EUA. 1994.

# Capítulo 10

## Noción de normalidad en Medicina: usos y limitaciones

*Mario Enrique Rendón Macías*



### Introducción

En este capítulo se analizan los dos enfoques conceptuales usados para determinar la normalidad. El primero de ellos considera la normalidad como la condición más usual o frecuente, asumiendo esta uniformidad como consecuencia de una evolución de nuestra especie.

Los valores más frecuentes se deben a la mejor adaptación de los individuos. Lo anormal es la variación más extrema en las características clínicas o los valores de una medición y en general resultado de adaptaciones muy particulares o variaciones momentáneas. Lo “normal” entonces se determina a través de límites de un espectro que por su frecuencia son los más esperados o usuales en una población de referencia. Se puede decir que esta es una definición estadística, aislada o univariable. Para determinar estos límites se ha recurrido a modelos matemáticos de probabilidad, como es la distribución de Gauss. El objetivo principal de esta definición es establecer y alertar a los médicos sobre posibles condiciones muy poco frecuentes que traducen variaciones sugestivas de inadaptación al medio o respuestas extremas al mismo.

El segundo enfoque tiene que ver con la decisión de actuar del médico. Este enfoque se denomina **definición correlacionada (multivariada) de normalidad**. Lo anormal se presenta cuando un dato o valor en el paciente implica la presencia de un comportamiento asociado a un proceso mórbido, un dato que surge como respuesta del organismo a esta enfermedad o bien a una condición perenne que facilita el inicio de una enfermedad.

De esta forma, una persona es normal mientras no esté incubando o padeciendo una enfermedad definida según criterios biológicos, cuando exista una acción terapéutica que mejore la salud o cuando se detecten condiciones clínicas modificables que eviten el surgimiento de una enfermedad.



En resumen, la normalidad puede ser enfocada en dos conceptos, uno relacionado con la frecuencia (univariada) con un objetivo de identificar condiciones extremas sugestivas de mala adaptación o respuesta extrema ante estímulos nocivos. La segunda, multivariada, correlacionada con procesos mórbidos o riesgos inherentes a su aparición o con la toma de decisiones terapéuticas.

## Caso clínico

---

En la sala de cuneros es revisado un recién nacido de 2 horas de vida, femenino, con un peso de 3 kg y talla de 50 cm. El neonato es hijo de la primera gestación de una mujer de 30 años de edad cuyo embarazo evolucionó sin ningún problema hasta sus 40 semanas. La madre acudió a la tocoquirúrgica por presencia de trabajo de parto con siete horas de evolución.

Al ser revisada por el médico de urgencias, detectó salida de líquido amniótico transvaginal claro, escaso. Al interrogatorio, la madre informó que el líquido empezó a salirle desde hace 8 horas. La niña nació por vía vaginal en un parto eutócico, cinco horas después del ingreso de la madre. Se le calificó de acuerdo a la escala de Apgar con un puntaje de 8 al minuto y 9 a los cinco minutos de nacer. No requirió de maniobras de reanimación. A la exploración de la niña se halla quejido espiratorio irregular, con frecuencia de 60 respiraciones por minuto y frecuencia cardíaca de 120 latidos por minuto. La temperatura corporal se registra en 38.1°C. Se ausculta una entrada de aire adecuada en ambos pulmones. Los reflejos primarios están presentes y la niña fija la mirada al estímulo luminoso.

A pesar del estado aceptable de la menor, se considera una posible infección por el antecedente de rotura prolongada de las membranas placentarias antes de su nacimiento. Como parte de una investigación en el hospital para buscar un marcador temprano de sepsis, se toman niveles de interleucina 6 (IL-6) y proteína C reactiva (PCR) de la sangre de la menor. A través de una prueba rápida se observa que el resultado fue de IL-6 = 245 pg/ml y para la PCR = 32 mg/L.

Ante estos datos las preguntas clínicas que surgen son: ¿los datos encontrados en la menor son normales o anormales?; ¿traducen una enfermedad?; ¿estos datos justifican realizar acciones terapéuticas?

---

Una de las tareas fundamentales del médico es establecer la condición de normalidad o anormalidad de un paciente. De ello depende la toma de decisiones para asegurar, corregir o mejorarle su salud. A pesar de ser un término común, su definición no es única y depende de los objetivos a resolver en el paciente.

## Concepto aislado de normalidad

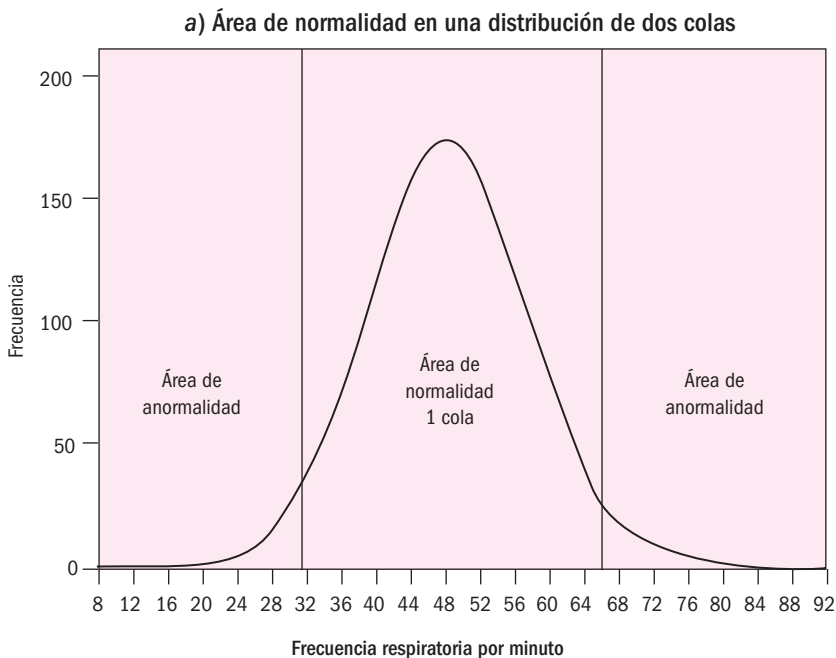
La definición de lo normal desde el punto conceptual de lo más usual o frecuente se fundamenta en que las variaciones encontradas al medir un atributo o característica en los grupos humanos no suelen ser muy grandes. La evolución de la especie humana con relación a su entorno ha condicionado que el comportamiento sea muy semejante entre uno y otro individuos, es decir, la frecuencia con la que las personas respiran, la

temperatura corporal, las horas de sueño, entre otros datos, aunque varían durante el día y periodos, en general tienden a mantenerse en unas cifras constantes. La pregunta que surge es ¿hasta dónde considerar estas variaciones como normales?

En cuanto al caso clínico que aparece en el recuadro anterior, el médico podría preguntarse: “¿cuán poco frecuente es encontrar en una recién nacida una cifra de 60 respiraciones por minuto o más?” Lo mismo pudiera decirse sobre una temperatura  $\geq 38.1^{\circ}\text{C}$  y una frecuencia cardiaca  $\geq 120$  por minuto.

Al considerar cada una de estas mediciones en forma separada (univariada), una respuesta para delimitar lo más frecuente fue la que propuso sir Ronald Fisher. Esta propuesta surgió al analizar el comportamiento de algunas mediciones, como la temperatura corporal, en una gran cantidad de sujetos, o en grupos de recién nacidos menores de siete días de vida, tal como se muestra en la figura 10-1a.

Esta distribución mostró varias características interesantes; la primera es que en general la mayoría de las cifras encontradas se agrupaba al centro de la distribución y la forma de la distribución semejava una campana por la simetría en relación con una cifra central. Más importante fue determinar que esta distribución podría ser obtenida matemáticamente con sólo una muestra aleatoria de la población; con ella se obtendría el valor medio sobre el cual la distribución mantiene su equilibrio de datos (media) y el comportamiento de



El área de normalidad incorpora al 95% de las observaciones;  
las colas en ambos lados es de 2.5% cada una.

**Figura 10-1.** Distribución normal según curva de Gauss. (Continúa)

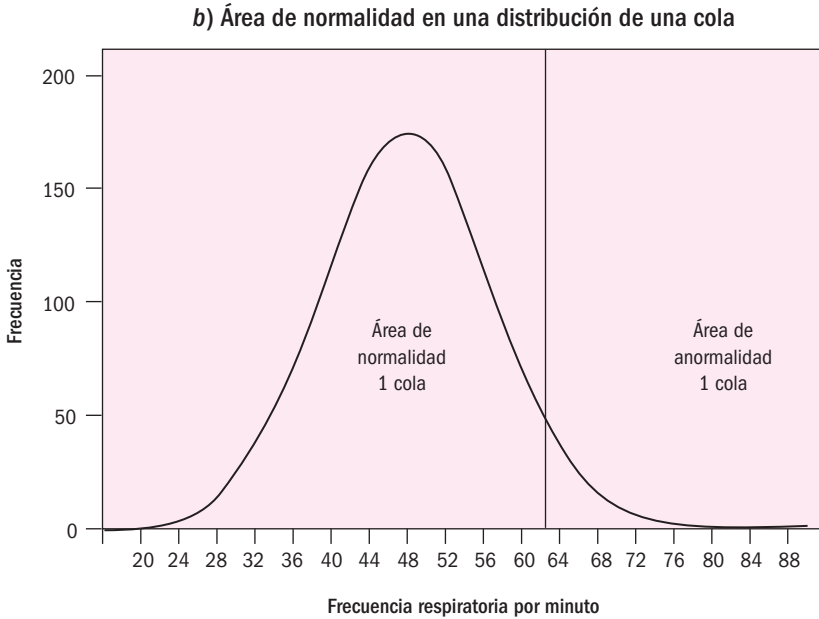


**Figura 10-1.** Distribución normal según curva de Gauss. (Continuación)

la frecuencia de los datos arriba y debajo de éste (dispersión). Esta distribución se conoce como “normal o gaussiana”. Para contestar a la pregunta sobre “¿cuán frecuente debe ser un dato para ser considerado como común o raro?”, los estadísticos estimaron la proporción de sujetos en una población, esperados con un dato particular. A fin de contestar a esta pregunta se observó que la curva normal tenía un punto en donde cambiaba de sentido (deflexión). A la distancia entre este punto y el valor de la media se le denominó una **desviación estándar**. Asumiendo que por debajo de la curva se encuentra toda la población, en esta área (figura 10-1*b*), le corresponde 33% de la población, tanto por arriba como por debajo de la media. Así que 33% de esta población tiene una frecuencia respiratoria entre 48 a 57 veces por minuto, o que si en una muestra de recién nacidos se tomara varias veces su frecuencia respiratoria, en alrededor de 33% de las veces estaría entre estas cifras. Esta última aseveración se conoce como el fenómeno de “regresión a la media” o “ley de los errores”, que establece que si un mismo objeto se mide de manera repetida, la distribución de estas mediciones se aproximará a la gaussiana o normal.

Con ayuda del valor promedio o media de la distribución y el valor de la distribución estándar, es posible estimar el valor “z” de un dato, que es la probabilidad de encontrar esta cifra en una población. Si uno desea saber “qué tan frecuente es encontrar en una recién nacida de 2 horas de vida con una frecuencia respiratoria de  $60 \times \text{min}$ ”, debe utilizar la fórmula siguiente:

$$z = \frac{X - \bar{X}}{DS}$$

Así:

$$z = \frac{60 - \overline{48}}{9.1} = 1.31$$

Con este valor se consultan las tablas de probabilidad en una curva normal en los libros de Estadística. De acuerdo con estas tablas, la probabilidad de encontrar esta frecuencia respiratoria o una mayor es de 9% ( $p = 0.09$ ). Debido al hecho de que en muchas mediciones las variaciones pueden ser amplias y muy extremas en algunas ocasiones, se tomó la decisión “arbitraria” de considerar como datos “anormales” a las cifras con menos de 5% de frecuencia o una proporción de 0.05 (probabilidad) de encontrarse; por tanto, según el criterio estadístico, este dato en la paciente puede ser una variación dentro de lo normal.

Aunque arbitraria, no es totalmente injustificada la lógica de pensar que una cifra tan extrema podría traducir una adaptación del organismo a un ambiente agresivo en el ser humano y, por tanto, compensatoria para mantener la homeostasis general. Una segunda opción es una conducta aberrante por un daño o alteración en el organismo. Estos datos, por tanto, deben alertar al médico sobre una alta probabilidad de una condición “anormal” o “algo anda mal”, más que una variación esporádica sin repercusión en la salud.

Existen dos formas de establecer la o las cifras “límite(s)” para definir lo “normal” o “usual” diferenciándolo de lo “anormal” o “de alerta”. Cuando en un dato se acepta como zona de alerta sólo los datos extremos hacia una dirección de la distribución (una cola) o hacia ambos lados de la distribución (2 colas), los valores considerados normales para la frecuencia respiratoria podrían ser:

- a) Una cola: normal, frecuencia respiratoria  $\leq 63$  por minuto
- b) Dos colas: normal, frecuencia respiratoria de 30 a 66 por minuto

Con estos criterios, para esta variable (frecuencia respiratoria) nuestro paciente se consideraría como normal y con un valor, aunque no frecuente, no tan extremo como para alertar a otra decisión médica diferente a la vigilancia habitual. Dicho criterio de normalidad estadística tiene las siguientes características.

Ventajas:

1. Establece la posibilidad de que existan variaciones observadas en la clínica al medir un dato como un fenómeno biológico.
2. Se dispone de cifras o límites claros para determinar qué tan extremos son los datos observados como para alentar al clínico sobre una posible enfermedad, tanto si existen síntomas clínicos como si no (cifras extremas en una determinación de glucemia sérica) y, aún más, avisar sobre el riesgo de padecerla (un nivel de colesterol extremadamente alto).
3. Tener certeza sobre un rango de valores encontrados en 95% de la población como un fenómeno de variación al azar.

## Desventajas:

1. Los valores frecuentes o normales pueden cambiar de una población a otra por adaptación de las mismas a otras condiciones ambientales o fisiológicas. ¿Sería el mismo comportamiento en la distribución de los valores de hemoglobina en poblaciones a diferentes altitudes, o en mujeres gestantes, o en prematuros? Es evidente que la evidencia científica ha mostrado que no es así. De ello se deriva la necesidad de ajustar valores de normalidad según grupos específicos; tal es el caso de la frecuencia respiratoria de nuestro paciente, al considerar a un recién nacido como un ser con un metabolismo diferente al de un escolar o un adulto.
2. Este criterio pudiera enmascarar a personas realmente enfermas. ¿Qué sucedería si los valores fueran estimados en una población con neumonía neonatal? Los datos más frecuentes en ellos no serían los más frecuentes en aquellos sin esta condición mórbida. Por ello es necesario establecerlo en personas “sanas o normales” bajo otro criterio.
3. Este criterio considera que lo normal es tener una cifra dentro de los valores encontrados en 95% de la población, lo cual implicaría que la anormalidad se encuentra siempre en el 5% restante, situación que no es real. Hay enfermedades cuya prevalencia es mucho mayor (la obesidad en México se ubica por arriba de 35%) o extremadamente baja (hipotiroidismo congénito  $<0.1\%$ ) bajo otros criterios.
4. Al determinarse la normalidad de forma univariada, si a un paciente se le realizan varias pruebas independientes, la probabilidad de que todas estuvieran en intervalos normales cada vez sería menor. Para una prueba la probabilidad de normalidad sería del 0.95, para dos de 0.9 ( $0.95 \times 0.95$ ) y para 20 sería de 0.35 o 35% de probabilidad de salir normal en todas. Es decir, “todos somos anormales” o diferentes en algo.
5. Este criterio se aplica si un dato tiene comportamiento en su distribución en la población de acuerdo con el modelo teórico de Gauss, pero, ¿y si no es así? Cuando un intervalo de valores de  $\pm 2$  desviaciones estándar no contiene al 95% de la población y/o en una distribución, la desviación estándar es mayor a la media en donde se generarán límites negativos no compatibles en muchas ocasiones con la clínica. Así, por ejemplo, la distribución de un dato como IL-6 con una media de 500 pg/100 ml y una desviación estándar de 420, daría límites de  $-340$  pg/100 ml y 940 pg/100 ml, la primera no compatible con la Biología. Esto se explica a continuación.

Existen datos clínicos que no se distribuyen de acuerdo con el modelo gaussiano, porque la mayoría de los datos se acumulan en un extremo; tal es el caso de variables como la frecuencia de embarazos en una población (figura 10-2).

En estos datos se observa que la mayoría de los pacientes tiene valores de cero o cercanos a ellos. La presencia de pocos o incluso un dato extremo puede afectar significativamente el valor de la media y por tanto distorsionar el cálculo de las desviaciones estándar. Una opción matemática es la transformación de los datos (logaritmos); sin embargo, esto no tiene sentido en su interpretación clínica. Así que una opción adecuada es trabajar con percentiles. Para ello se hace caso omiso a la forma de distribución de los

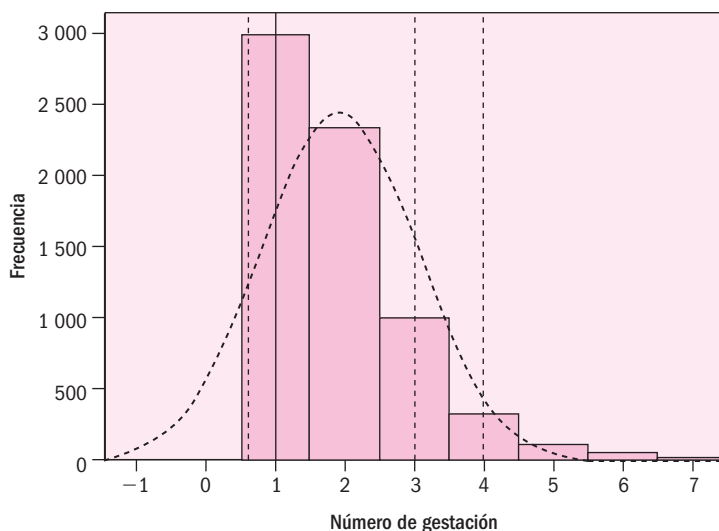


**Figura 10-2.** Distribución de número de gestaciones.

datos y se organizan por valores jerárquicos en orden ascendente de menor a mayor, luego se localiza la zona central (95%) de los valores observados y se buscan los percentiles 2.5 (que limita 2.5% de los valores en el extremo inferior) y el 97.5 (que limita 2.5% de valores en el extremo superior). Al igual que los límites en la distribución normal, cifras por debajo del percentil 2.5 o arriba del 97.5 se consideran muy raras y, por tanto, motivo de alertar al médico. Esta estrategia corrige el problema de la no distribución gaussiana de los datos, pero tiene las mismas implicaciones o desventajas de ésta.

## Concepto relacionado de normalidad

El concepto anterior dejó claro que la variabilidad de los datos clínicos se debe a cambios en la medición (fenómeno del azar) o cambios debido a compensaciones del organismo a factores ambientales o nuevas adaptaciones a consecuencia de un daño. Estas dos últimas consideraciones han generado uno de los conceptos actuales de enfermedad. Considere cómo la relación de varios acontecimientos o condiciones puede permitir definir un estado de normalidad o anormalidad. En el concepto relacionado de normalidad, existen tres criterios a tomar en cuenta: *a)* presencia de síntomas y/o signos, *b)* una exposición a un factor de riesgo y *c)* disponibilidad de una terapéutica. La importancia de este concepto es la posibilidad de la toma de decisiones para el actuar del médico.

### A. Presencia de síntomas y/o signos

La Medicina surgió como la respuesta de una serie de personas (futuros médicos) para mejorar las condiciones de salud de otras (pacientes), quienes manifestaban padecer o

sentir algo raro o distinto en su cuerpo. El estudio de grupos de individuos con padecimientos iguales o semejantes, y su comparación con el comportamiento del resto de la población sin estas dolencias, fue lo que definió el concepto de una enfermedad. Más adelante, el estudio de la Fisiología y Fisiopatología, aunado a otras ciencias, ha definido cada vez con mayor precisión la forma de agrupar y nombrar a estos pacientes. De esta forma, se considera anormal a una persona que presenta síntomas o signos que no le son habituales, pero que sí son frecuentes en entidades nosológicas o enfermedades bien reconocidas.

En el caso clínico señalado en el recuadro se hace referencia a una niña recién nacida con un dato que se considera poco frecuente en la población de esta edad, “quejido espiratorio irregular”. Ya que es un dato observado, se considera un signo. Si se realizara un estudio para buscar cuántos neonatos de esta edad presentan este dato, se encontraría una probabilidad (proporción) alrededor de 1%, lo cual traduce un dato poco frecuente y quizá ni siquiera motivo de alarma. Sin embargo, la actitud del médico cambiará de manera drástica si considera que presentan un quejido espiratorio hasta 50% de los pacientes diagnosticados con neumonía intrauterina por tener evidencia de alguna bacteria en su sangre, con invasión de su tejido pulmonar y una respuesta inflamatoria y alto riesgo de fallecer.

De esta forma, se considera anormal a una persona que manifiesta síntomas o signos asociados o relacionados con una enfermedad. Diferentes estudios han sido llevados a cabo durante años para determinar los datos que corresponden a cada enfermedad o los más frecuentes. Pero ha sido la combinación de éstos, reunidos bajo comportamientos fisiopatológicos comunes, lo que ha definido la condición de normalidad o no, es decir, si se tienen los datos asociados a una enfermedad, se es anormal.

Regresando al caso clínico citado, es la combinación de datos clínicos, como quejido espiratorio, una frecuencia respiratoria con una cifra poco común y una temperatura también registrada en un valor poco común, lo que haría sospechar una neumonía temprana.

Debido a la necesidad de disponer de mayor evidencia para documentar una enfermedad o condición anormal, los estudios de laboratorio e imagenológicos han buscado ampliar nuestros sentidos en la observación de las condiciones reales de los pacientes. En esta paciente se tomaron niveles de IL-6 y PCR. Ambas sustancias no son visibles sin el recurso de un laboratorio. Con el conocimiento básico y clínico previo, se ha determinado la distribución de los niveles en grupos de pacientes con los criterios de neumonía y sepsis neonatal contra aquellos de neonatos sin ninguna evidencia de infección bacteriana. Los resultados mostraron que en niños con sepsis o neumonía temprana los niveles de IL-6 variaron de 50 a 10 000 pg/100 ml *vs.* de 0 a 95 pg/100 ml en los controles. Encontrar cifras por arriba del criterio univariado de anormalidad es considerado como otro dato clínico que con los otros traduce alta probabilidad de pertenecer al grupo de los neonatos con sepsis y por lo tanto ser anormal o enfermo.

En resumen, se considera a una persona como anormal cuando tiene síntomas o signos correspondientes a una entidad nosológica reconocida. Para definirla es necesario que cumpla con los criterios ya establecidos para la enfermedad en cuestión.

## Ventajas:

1. El criterio de anormalidad se sustenta en aquellos con una enfermedad definida, de tal forma que se respetan las prevalencias reales de las enfermedades.
2. Los criterios no son arbitrarios en relación con un punto de vista estadístico.
3. Un solo dato extremo puede ser tomado con prudencia, dada la posibilidad de un error en el procesamiento del dato.
4. Por ser un criterio establecido, permite al médico agrupar al paciente y determinar una conducta terapéutica, pronóstica o de mayor escrutinio diagnóstico.

## Desventajas:

1. En ocasiones una enfermedad puede estar en su fase subclínica o inicial con pocos datos que no permiten completar un criterio establecido de diagnóstico. La decisión sobre el inicio de un tratamiento puede ser difícil, más si existen efectos secundarios graves o frecuentes a éste. No conseguir determinar un estado como anormal quizá conlleve no iniciar un tratamiento en una fase oportuna o quizá la única posibilidad de éxito.
2. Hay padecimientos manifestados por algunos pacientes que no han sido catalogados como una enfermedad, lo cual deja al médico en una situación difícil para decir si es anormal. Además, no es de extrañar la existencia de enfermedades actualmente desconocidas y, por tanto, no diagnosticables, lo cual no asegura que el paciente está sano.
3. En ocasiones existen condiciones que pueden ser catalogadas como anormales y por tanto enfermedades; sin embargo, no es prudente emitir las por las consecuencias sociales existentes, tal como sucede en personas con alguna mutación conocida detectada al nacimiento, por ejemplo, la corea de Huntington. El individuo no manifestará síntomas hasta llegar a su vida adulta; mientras tanto, puede ser considerado como normal o sano.
4. Existen enfermedades en donde no se cuenta con tratamientos eficientes y el comportamiento puede ser estable, por lo cual emitir una condición de anormalidad pone al médico en una situación difícil para decidir su acción. Tal es el ejemplo de alguna anormalidad congénita benigna como una costilla de más. Lo consideraría anormal si el paciente aqueja dolor o molestia por la misma, pero normal si esta condición es asintomática.

## B. Exposición a un factor de riesgo

Decidir esperar a que un paciente complete los datos clínicos necesarios para definir una enfermedad puede tener consecuencias desastrosas. Si la recién nacida del caso clínico tiene una bacteria en su sangre, esperar a que manifieste todos los síntomas definidos como una neumonía pudiera ser peligroso, dado que algunos neonatos con ella pueden pasar de pocos síntomas a una sepsis grave con choque séptico en cuestión de horas. Por ello, los médicos han explorado indicadores o marcadores preclínicos para detectar



a estos pacientes en riesgo y dar un tratamiento aún más oportuno; es decir, un tratamiento que elimine al agente causal antes de que comience a lesionar al cuerpo. De esta forma, detectar un factor que se asocia fuertemente con una enfermedad, hace al sujeto como “anormal o en riesgo”. El factor puede ser detectado por medio de estudios epidemiológicos previos; tal es el caso de la rotura prematura de membranas (más de 12 h) como condición que aumenta la probabilidad de presentar sepsis temprana, aun cuando un neonato no tenga ningún síntoma. Esta condición incrementa el riesgo hasta en cinco veces más que no tener el antecedente.

Cuando el riesgo es muy alto y existe un tratamiento efectivo para controlar la posible enfermedad, la decisión de administrarlo se justifica y, por tanto, no es necesario esperar a la aparición de síntomas para definir a la persona como anormal. Debido a lo anterior, la decisión de iniciar la administración empírica de antimicrobianos en esta paciente pudiera estar justificada, hasta la confirmación o no del aislamiento bacteriano.

En resumen, una persona se considera anormal cuando se demuestra la presencia de un factor de riesgo altamente asociado con una enfermedad.

#### Ventajas:

1. La anormalidad puede ser buscada y no sólo considerada cuando un paciente manifiesta un padecimiento o es detectado durante una revisión.
2. Conocer factores de riesgo ha justificado iniciar manejos profilácticos y reducir la incidencia de muchas enfermedades.
3. Como criterio correlacionado, se basa en evidencia epidemiológica y no en criterios estadísticos. De ahí que los criterios no son arbitrarios.

#### Desventajas:

1. La principal desventaja es que toda persona está expuesta a factores de riesgo asociados a enfermedades; por tanto, todos seríamos anormales. Así, por ejemplo, y citando el caso clínico mencionado, además de considerarse anormal por riesgo de infección temprana por la rotura prematura de membranas, también podría decirse que es anormal por riesgos distintos, como sitio de nacimiento, tipo de atención, etcétera. Estos riesgos no se asocian a sepsis pero sí a otras enfermedades o condiciones clínicas futuras, como desarrollo psicomotor, educación, entre otros.
2. Existen factores de riesgo conocidos para algunas enfermedades, pero sin conocerse tratamientos efectivos para su eliminación. Determinar a una persona como “anormal” por su riesgo sin ofrecerle una opción no es éticamente aceptado. Esto ha disminuido el interés en conocer el genoma de las personas.
3. Afirmar que un riesgo es sustancial es subjetivo, ya que aunque se acepta como riesgos altos los mayores de dos (razones de momio, riesgo relativo u otros), éstos dependen en gran medida del grupo control sobre el cual se llevó a cabo el cálculo.
4. En algunas enfermedades se desconocen los factores de riesgo y, por tanto, sólo se puede actuar cuando éstas presentan manifestaciones clínicas.

### C. Disponibilidad de una terapéutica

El objetivo principal de asignar un nombre o diagnóstico a la condición clínica de un paciente es tener la mayor certeza en las posibles consecuencias de nuestras decisiones terapéuticas. Bajo este criterio se considera a una persona anormal o enferma en el momento en que dar una maniobra profiláctica o terapéutica empieza a ser más benéfica que no darla.

Como ya se comentó, en el contexto del caso clínico señalado, la existencia de una serie de maniobras terapéuticas —como ayuno, registro constante de los signos vitales, el inicio de un antimicrobiano de amplio espectro— representa acciones con menor riesgo de daño que el beneficio sustancial de controlar una infección potencialmente grave. El médico tomaría la decisión de definirlo como anormal y justificar sus acciones.

#### Ventajas:

1. La decisión de una condición normal y anormal va relacionada directamente con ofrecer una opción eficaz para mejorar la condición de salud del paciente.
2. Se puede clasificar a una persona como anormal tanto en fases preclínicas como clínicas e incluso con riesgos conocidos, con lo que se busca un mejor impacto sobre la salud de las personas.
3. Condiciones clínicas particulares pueden ser consideradas como normales sin la repercusión social que conllevan; tal es el caso para niños con dolicocefalia, a quienes, aun siendo diferentes, se les trata como sanos.
4. El criterio de anormalidad es claro y basado en evidencias científicas, con altas probabilidades de justificarse.
5. En ocasiones una respuesta a una maniobra terapéutica puede confirmar la presencia de una enfermedad no conocida o no detectable con otras pruebas. Su evidencia puede iniciar la definición de una nueva enfermedad y la detección de otros pacientes. La respuesta al uso de esteroides en algunos pacientes con síndrome nefrótico generó una nueva clasificación.

#### Desventajas:

1. Considerar a una persona anormal sólo cuando existe un tratamiento eficaz para su control implica considerar normales a pacientes con enfermedades incurables.
2. Con el avance de la ciencia y el incremento de las normas éticas en la atención de pacientes, se pugna cada vez más por otorgar manejos curativos o paliativos a todo paciente; si no existe curación, se deben realizar acciones para mejorar la calidad de vida. De esta forma, de nuevo toda persona amerita algún manejo para mejorar su salud.
3. Se puede justificar una condición anormal ante la posibilidad de una respuesta a una maniobra en fase de experimentación o con poco sustento clínico epidemiológico. En ocasiones esto puede dar origen a tratamientos cuestionables, como cirugías plásticas.
4. La definición de un tratamiento eficaz depende del criterio elegido para definirlo y la comparación realizada contra el tratamiento control o su ausencia. De esta forma, existen muchos tratamientos para diferentes enfermedades o condiciones clínicas.

## Conclusión

Definir a una persona como **normal** o **anormal** depende del criterio considerado. Cualquier criterio tiene ventajas y desventajas, y entre ellas es factible existan contradicciones. Por lo anterior, la decisión debe ser guiada por los objetivos en la atención de un paciente buscando el mayor beneficio para éste. Por ello, es necesario, cuando se define la condición de anormalidad de un paciente, analizar el o los criterios utilizados para su enunciación.

En el cuadro 10-1 se resumen las características, ventajas y desventajas de cada criterio.

En el caso mostrado, el criterio univariable puso a la paciente en una condición de normalidad por sus signos clínicos, normal en cuanto a una enfermedad conocida (sepsis neonatal o neumonía temprana) por no completar los criterios establecidos; pero como anormal ante el riesgo de una sepsis potencialmente grave (antecedentes de rotura prematura, quejido espiratorio y una elevación de IL-6) y anormal por el potencial beneficio de un manejo temprano con antimicrobianos ante la sospecha diagnóstica.

**Cuadro 10-1.** Comparación de los criterios de normalidad

| Criterio                              | Definición                                                              | Ventaja                                                                                                                                                                                                                                                                               | Desventaja                                                                                                                                                                                           | Aplicación                                                                                                |
|---------------------------------------|-------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Univariado con distribución gaussiana | Valores encontrados entre $-2$ y $+2$ desviaciones estándar de la media | Considera lo más frecuente.<br>Delimita límites claros.<br>Puede detectar problemas en etapas subclínicas o datos de riesgo                                                                                                                                                           | Se basa en modelos estadísticos.<br>A más estudios mayor probabilidad de un dato anormal.<br>No todos los datos clínicos muestran una distribución normal                                            | Valores extremos alertan sobre posibles enfermedades                                                      |
| Univariado sin distribución gaussiana | Valores comprendidos entre los percentiles 2.5 y 97.5                   | Para datos sin distribución normal.<br>Igual al anterior                                                                                                                                                                                                                              | Igual al anterior                                                                                                                                                                                    | Valores extremos alertan sobre posibles enfermedades                                                      |
| Correlacional por sintomatología      | Se es anormal si existen datos clínicos observables o medibles          | Respetan la prevalencia real de las enfermedades.<br>Se asocian a condiciones fisiopatológicas establecidas.<br>Permiten tomar decisiones con base en el conocimiento de la enfermedad, tales como establecer pronósticos.<br>Los criterios tienen bases biológicas y no estadísticas | No considera las fases de latencia y preclínicas.<br>El definirlas con este criterio puede ser en un momento tardío cuando un tratamiento ya es ineficiente.<br>No considera enfermedades emergentes | Clasifica a las personas en condiciones nosológicas establecidas y permite determinar acciones a realizar |

(continúa)

**Cuadro 10-1.** Comparación de los criterios de normalidad (Continuación)

| Criterio                             | Definición                                                                                   | Ventaja                                                                                                                                                                                                      | Desventaja                                                                                                                                                                                                      | Aplicación                                               |
|--------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| Correlacional por riesgo             | Se es anormal cuando se tiene un factor de riesgo asociado con una entidad establecida       | Permite la toma de decisiones terapéuticas tanto en sujetos enfermos (en fases clínicas y preclínicas) como sanos con un riesgo conocido. Permite realizar acciones oportunas en tiempo                      | Todas las personas están expuestas a un riesgo. El valor de un riesgo depende de la comparación a la que se calculó. Riesgos no modificables injustifican su determinación                                      | Detección de personas para tratamientos tempranos        |
| Correlacional por tratamiento eficaz | Se es anormal cuando se dispone de un tratamiento más benéfico que perjudicial para la salud | Permite tomar decisiones terapéuticas en cualquier condición clínica del paciente. Delimita criterios basados en evidencia científica. Sólo se consideran riesgos con tratamientos eficaces para eliminarlos | Las enfermedades incurables generan dificultad en su clasificación. Toda condición es susceptible de ser tratada por algún método. La eficacia se basa en la comparación realizada y los objetivos considerados | Permite tomar decisiones de actuación directa del médico |

## Bibliografía

- Beceiro-Mosquera J, Sierva-Monzo CL, Oria de Rueda-Salguero O, Olivas-López de Soria C, Herbozo Nory C. Utilidad de un test rápido de interleuquina-6 en recién nacidos con sospecha de infección. *An Pediatr (Bar)* 7(16):483-488. 2009.
- Clifford Blair R, Taylor RA. *Bioestadística*. Pearson, Prentice Hall, México, 415-419. 2009.
- Fletcher RH, Fletcher SE. *Epidemiología clínica*, 4a. ed. Wolters Kluwer, Lippincott/Williams & Wilkins editores, Barcelona, España, 19-37. 2007.
- Gómez-López N, Laresgoiti-Servijte E, Olson DM, Estrada-Gutiérrez G, Vadillo-Ortega F. *The role of chemokines in term and premature rupture of the fetal membranes: a review*. BOR paper in press. Published on January 20, 2010 as DOI:10.1095/biolreprod.109.080432.
- Idriayan A. Medical Biostatistics. Chapman & Hall/CRC, Taylor & Francis Group (editores). 2a. ed. EUA, 233-240. 2008.
- Rusconi F, Castagneto M, Gagliardi L. Reference values for respiratory rate in the first 3 years of life. *Pediatrics* 94:350-355. 1994.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Epidemiología Clínica. Ciencia básica para la medicina clínica*, 2a. ed. Editorial Médica Panamericana, Argentina, 19-78. 1994.
- Song-Ro Y, Dubeau L, de Young M, Wexler NS, Arnheim. Huntington disease expansion in humans can occur before meiosis is completed. *PNAS* 100(15):88. 2003.



# Capítulo 11

## Concepto de causalidad en Medicina

*Florencia Vargas Voráckova (q.e.p.d.)*

### Introducción

Una asociación puede definirse como la dependencia estadística que existe entre dos o más factores, donde la ocurrencia de un factor aumenta (o disminuye) a medida que varía otro. Así, por ejemplo, la frecuencia de la hipertensión arterial aumenta conforme se incrementa la edad de la población.

La presencia de asociación no implica necesariamente causalidad. Una asociación causal (o relación de causa-efecto) se manifiesta cuando el cambio en la frecuencia o intensidad de la exposición es seguido por un cambio en la frecuencia o intensidad del efecto. La exposición al bacilo de Koch y el desarrollo subsecuente de tuberculosis, o bien la administración de penicilina seguida de la curación de una faringoamigdalitis estreptocócica, son ejemplos de asociación causal.

Las causas pueden ser de dos tipos, necesarias y suficientes. Los factores indispensables para el desarrollo de efectos, es decir, aquellos sin los cuales los efectos nunca podrían aparecer, se conocen como causas necesarias. Así, en ausencia del bacilo de Koch no puede haber tuberculosis. Por otra parte, aquellos factores que, en conjunto (incluyendo la causa necesaria), siempre culminan con el desarrollo del efecto se denominan **causas suficientes**. En condiciones de higiene y buen estado nutricional, una persona sana expuesta al bacilo de Koch difícilmente desarrolla tuberculosis; sin embargo, un sujeto que vive en situación de hacinamiento, malnutrición e inmunocompromiso tiene altas posibilidades de presentar la entidad. En este último caso, el bacilo de Koch es la causa necesaria que, aunada a hacinamiento, malnutrición y compromiso inmunitario, conforma el grupo de causas suficientes para el desarrollo de la enfermedad. En otra terminología, causa es equivalente a causa necesaria, y factores de riesgo, a causa suficiente (excepto la necesaria). Al suprimir la causa, el efecto se erradica; al eliminar un factor de riesgo, el efecto disminuye en frecuencia.

## Caso clínico

---

Una mujer primigesta, de 28 años de edad, sin antecedentes de importancia, acude al médico con motivo de su primera visita de control de embarazo. La paciente refiere que su esposo es un gran fumador y pregunta si esto puede influir sobre el desarrollo adecuado de su producto.

La relación entre tabaquismo pasivo y producto de bajo peso al nacer es familiar al médico y recuerda el trabajo clásico que Martin y Bracken efectuaron en 1982. Dicho estudio se realizó con el fin de evaluar los efectos del tabaquismo pasivo de la madre sobre el peso del producto al nacer y sobre otras variables de crecimiento fetal. Para ello, estos autores entrevistaron a 4 186 mujeres (de un total de 6 219) que acudieron a los servicios de salud del área de New Haven, Estados Unidos, con motivo de su primera evaluación prenatal. El cuestionario abordó información sobre antecedentes ginecoobstétricos, características demográficas, métodos anticonceptivos, historia clínica y exposición a otros factores de riesgo. El tabaquismo pasivo se definió como la exposición de la mujer embarazada al humo expelido por un fumador (que no fuera ella misma) durante un periodo mínimo de 2 horas diarias. La información sobre el resultado del embarazo se obtuvo subsecuentemente de los expedientes clínicos respectivos. Cuando el peso del producto al nacer fue menor a 2 500 g, se consideró como peso bajo. Se obtuvo la información deseada en 3 858 productos únicos, de los cuales 3 610 nacieron a término. En 2 473 de estos últimos, la madre no fumó activamente durante el embarazo. Si se considera a las madres de productos a término y no fumadoras activas, 853 estuvieron expuestas pasivamente al humo del tabaco y 1 620, no. El porcentaje de productos con bajo peso al nacer fue de 2.34 en el primer grupo y de 0.86 en el segundo.

A partir de este estudio, ¿se puede afirmar que existe asociación entre tabaquismo pasivo y bajo peso al nacer? De ser así, ¿es esta una asociación causal?

Sin embargo, el médico sabe que mientras la relación entre tabaquismo materno y problemas como bajo peso al nacer está bien establecida, los efectos del tabaquismo pasivo han sido menos estudiados y los estudios realizados muestran poca consistencia. Así, decidió revisar la literatura actual, como el estudio que Salmas et al. realizaron en 2010 en el Departamento de Ciencias de la Salud de la Universidad de McMaster en Hamilton, Ontario, Canadá, el cual tenía por objetivo determinar el efecto del tabaquismo pasivo en recién nacidos. Los autores estudiaron 66 artículos de un total de 48 439 mujeres expuestas al humo de tabaco y 90 918 mujeres no expuestas.

---

## Importancia del conocimiento de la causalidad

Para el médico, el conocimiento de la causalidad es de gran utilidad en diversas circunstancias del acto médico. Se utiliza en el momento de decidir si el agente etiológico “X” produce la enfermedad “Y”; si el factor de riesgo “XX” favorece una mayor frecuencia; si el agente terapéutico “A” cura el padecimiento o evita que avance hasta no generar incapacidad, o bien si la medida profiláctica “B” previene su aparición.

Así, una vez que el bacilo de Koch fue identificado como agente causal de la tuberculosis, pudieron procesarse tanto la vacuna BCG como fármacos efectivos para el tratamiento de la infección (p. ej., isoniazida, rifampicina, etambutol). Por otro lado, la

identificación de factores de riesgo, como hacinamiento y desnutrición, ha permitido disponer de recomendaciones higienicodietéticas para reducir la frecuencia de la enfermedad.

Los efectos colaterales de ciertos medicamentos y la influencia que éstos ejercen sobre las indicaciones terapéuticas de un médico reflejan otras de las aplicaciones del conocimiento de la causalidad. De esta forma, se sabe que la cimetidina (un bloqueador de receptores  $H_2$  utilizado para el tratamiento de enfermedad acidopéptica) favorece la aparición de ginecomastia en pacientes de sexo masculino.

Si bien en la Medicina actual son muchos los padecimientos cuyas causas se conocen, mayor es el número de aquellos cuyas causas se ignoran. En este último grupo figura la mayoría de las enfermedades cronicodegenerativas (p. ej., hipertensión arterial, neoplasias malignas), para las que sólo ha sido posible identificar algunos factores de riesgo.

## Determinación de la causalidad

Para concluir que dos o más factores están causalmente relacionados se requiere demostrar que la asociación entre estos factores es válida y causal.

### Asociación válida

Una asociación real (o verdadera) es una asociación válida. Esta validez conlleva un efecto mínimo del azar (error aleatorio) o sesgo (error sistemático) o ambos (véase capítulo 9).

### Azar

El papel del azar en el hallazgo de una asociación puede ponderarse al aplicar una prueba de significación estadística (p. ej., chi [ji] cuadrada, exacta de Fisher) y expresarse en función del valor  $p$  resultante. Un valor  $p$  igual o menor que 0.05 es estadísticamente significativo y traduce una probabilidad atribuible al azar de 5% o menos. De aquí que una asociación con importancia estadística tenga altas probabilidades de ser real, dado que el efecto del azar se restringe a una probabilidad máxima de 5%.

Otro método para evaluar la influencia del azar es la determinación del intervalo de confianza a 95% (IC 95%). Este intervalo refleja la información que se esperaría obtener en 95 de 100 replicaciones del mismo estudio. Así, el IC 95% de una medida de asociación contiene, con 95% de seguridad, el valor real de esta medida.

Pese a que su propósito es el mismo, la información brindada por el valor  $p$  y el IC 95% es un tanto diferente. En términos del valor  $p$ , una asociación que no es estadísticamente significativa (es decir, con  $p > 0.05$ ) tiene pocas probabilidades de existir y, por tanto, puede ignorarse. Sin embargo, el intervalo de confianza quizá no excluya la posibilidad de que exista dicha asociación y, en forma adicional, es posible que proporcione cierta información sobre su magnitud. Es frecuente encontrar este tipo de situaciones, en especial en estudios pequeños, ya que tanto el valor  $p$  como el IC 95% son susceptibles al tamaño de la muestra: a menor tamaño de ésta, mayor es el valor  $p$  y más amplio el intervalo.

## Sesgo

La presencia de sesgo puede distorsionar la información de una asociación. Así, es posible que haga aparecer una asociación que en realidad no existe o la vuelva mayor de lo que en verdad es, o bien la torne menor. Los sesgos pueden ocurrir en cualquier fase del proceso de evaluación de una asociación: destacan los sesgos de selección, medición y confusión.

Por el proceso de selección, es posible que ciertos sujetos tengan mayor probabilidad de ser incluidos o permanecer en un estudio. El sesgo de selección sucede en estas circunstancias siempre y cuando los individuos seleccionados difieran en características importantes de la población de origen (a la que se pretende extrapolar los hallazgos del estudio) o de otro grupo de personas (con el que se busca hacer comparaciones). Ejemplo de este tipo de sesgos es el de Berkson, el cual se presenta en ciertos estudios de casos y controles realizados en hospitales, y se manifiesta con una probabilidad de inclusión variable en los diferentes grupos de efecto y exposición. Por ejemplo, si la probabilidad de ser incluido como caso o como expuesto es mayor, o ambas cosas, esto puede derivar en la sobrestimación de una asociación.

El sesgo de medición ocurre cuando durante la fase de obtención de información los investigadores utilizan, para cada uno de los grupos estudiados, criterios diferentes de recolección o interpretación de datos, o ambos. También puede presentarse cuando la información brindada por los sujetos en estudio no es comparable en los grupos analizados. Como ejemplo de las condiciones que favorecen que suceda este sesgo, considere el uso de pruebas de detección de un efecto, cuya sensibilidad y especificidad varían en función de la exposición. En este caso, el grado de sobrestimación o subestimación de la asociación depende del número de resultados falsos positivos y negativos determinado por cada prueba.

La asociación entre dos factores puede verse afectada por la coexistencia de otros, conocidos como factores de confusión. Para considerarlos como tales, los factores de confusión deben estar asociados a los factores de la asociación en estudio.

En una asociación causal, el factor de confusión ha de estar asociado tanto a la exposición como al efecto, y la asociación entre factor de confusión y efecto es de tipo causal (véase capítulo 9).

El efecto del azar y el sesgo en la evaluación de una asociación disminuye debido al rigor metodológico del estudio. Se describen a continuación tres diseños, y se mencionan brevemente otros, en orden de rigor metodológico decreciente.

El diseño que permite obtener la información más confiable es el de un experimento (véase capítulo 5), en el que los sujetos de una muestra, después de asignarse en forma aleatoria para que reciban o no cierta exposición, se observan durante cierto periodo, al cabo del cual se determina quiénes desarrollaron el efecto y quiénes no (figura 11-1).

En este caso, la asignación aleatoria permite obtener grupos similares o comparables. Este diseño, que también se conoce como **ensayo clínico aleatorio**, muy pocas veces puede llevarse a cabo en el proceso de investigación de causa o factores de riesgo (o am-



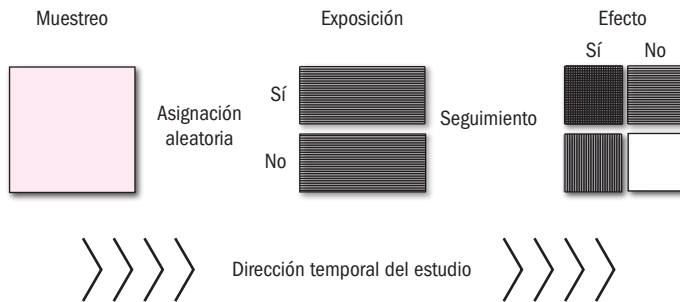


**Figura 11-1.** Diseño de un ensayo clínico aleatorio.

bos) de enfermedades. La razón de ello está en los aspectos éticos, en el número de sujetos por estudiar y en el tiempo de realización. No es ético imponer de manera aleatoria que una persona se exponga o no a un factor causal o de riesgo.

Por otra parte, si la aparición del efecto es baja, se requiere estudiar un gran número de casos expuestos y no expuestos, lo cual ocasiona reclutamiento prolongado o de altos costos, o ambos. Por último, si se considera el tiempo que se requiere seguir a los sujetos expuestos y no expuestos para observar cuántos desarrollan el efecto, este periodo debe incluir años, lo cual implica gran consumo, no sólo de tiempo, también de recursos humanos y monetarios. Asimismo, un seguimiento de años de duración presenta ciertas desventajas metodológicas que pueden afectar en un momento dado la validez de los resultados; entre éstas destacan el seguimiento irregular y la pérdida de sujetos en estudio, ambos debido a desmotivación, aburrimiento y otros factores. Simplemente, al cuestionar lo ocurrido a un paciente que se perdió durante la fase de seguimiento de un estudio, surgen varias posibilidades: ¿se perdió por haber presentado el efecto que interesa investigar o porque se cansó de hallarse en seguimiento? La validez del estudio se ve amenazada en función de la proporción de pacientes perdidos o cuyo seguimiento se efectúa en forma irregular.

Una alternativa metodológica, también comparativa y prospectiva pero que no recurre a la asignación aleatoria, es el estudio de cohortes. En este tipo de estudios, los sujetos de una muestra son clasificados según su estado de exposición en una cohorte expuesta y en otra no expuesta, y se observan de manera subsecuente durante un periodo determinado a fin de evaluar la frecuencia de aparición del efecto (figura 11-2). En este caso, son otras las circunstancias, diferentes a la asignación aleatoria, las que deciden que un individuo esté expuesto o no a la causa o al factor de riesgo, y aquellas que en determinado momento pueden permitir que ocurra sesgo.

En tal situación, existe la posibilidad de que la circunstancia determinante de la presencia o ausencia de la exposición en realidad constituya la causa o el riesgo del efecto buscado.

El estudio de cohortes tampoco garantiza una distribución similar de características entre los grupos expuesto y no expuesto (comparabilidad de grupos) que en el momento del análisis, permita descartar un efecto de confusión y atribuir el efecto observado a la ex-

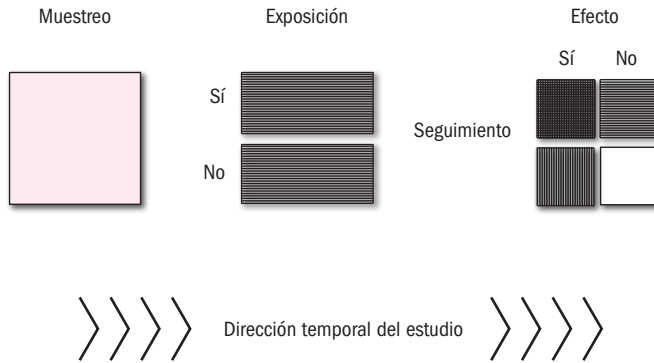


**Figura 11-2.** Diseño de un estudio de cohortes.

posición estudiada. Además de estos inconvenientes metodológicos, el diseño de cohortes comparte con el ensayo clínico aleatorio los problemas referentes al tamaño de muestra y al seguimiento.

Sin embargo, por las razones éticas descritas, es más factible de realizar que el ensayo aleatorio y, por tanto, es frecuente encontrarlo en la literatura médica.

El diseño comparativo más fácil y rápido de efectuar, aunque metodológicamente más débil, es el de casos y controles. En éste se parte de un grupo de casos (sujetos que ya presentan el efecto) y de un conjunto de testigos o controles (individuos que no muestran el efecto) para investigar en retrospectiva el antecedente (o la ausencia de éste) de exposición a la causa o al factor de riesgo (figura 11-3).

Este tipo de estudio tiene aún más posibilidad de que se comprometa la comparabilidad de los grupos. Además de lo mencionado en los estudios de cohorte, el hecho de partir de etapas posteriores en la secuencia de exposición y efecto, puede volver cuestionable la representatividad de los grupos estudiados. En otras palabras, ¿cuántos sujetos expuestos y no expuestos que desarrollaron o no la enfermedad se perdieron y, por tal motivo, no fueron incluidos en los grupos de casos y controles estudiados? ¿Qué tan

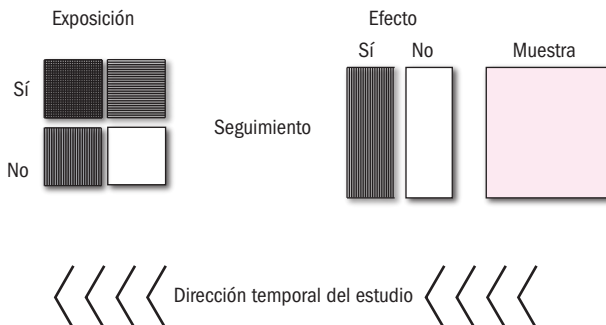


**Figura 11-3.** Diseño de un estudio de casos y controles.

representativos son los casos y controles del estudio respecto de los casos y controles originales? Una limitante adicional de este tipo de estudios es la dificultad de asegurar una secuencia temporal lógica, donde la exposición antecede al efecto. Este es el tipo de diseño que se encuentra con más frecuencia en los trabajos publicados sobre causas y factores de riesgo de enfermedades.

En el diseño transversal, que es una imagen instantánea de la coexistencia simultánea (en un punto determinado de tiempo) de la exposición y el efecto, y que comparte las mismas limitaciones del estudio de casos y controles, la dificultad para demostrar la secuencia temporal de exposición y efecto es más acentuada (véase capítulo 4).

Finalmente, las alternativas metodológicas menos recomendables son los diseños descriptivos. En ellos sólo se estudia en forma prospectiva un grupo de sujetos expuestos o, retrospectivamente, un conjunto de casos. Si no se tienen controles o grupos de referencia con los cuales comparar, ¿cómo se puede estar seguro de que la frecuencia del efecto es diferente en los no expuestos o que la exposición es distinta en un grupo que no presenta el efecto?

En determinado momento, la selección de un diseño en los estudios de causalidad depende de lo éticamente posible y de los recursos disponibles.

Debe procurarse seleccionar el mejor método para acercarse a la “verdad”, de manera que al contar con información válida, ésta resulte en beneficio de los enfermos y permita mejor aprovechamiento de recursos.

## Asociación causal

Una vez que se ha demostrado que una asociación es válida o real, el siguiente paso consiste en comprobar que esta asociación es causal. Son nueve los criterios propuestos por **Bradford-Hill**, mismos que se describen en el cuadro 11-1. La información requerida para satisfacer estos aspectos no se fundamenta sólo en la del estudio en cuestión, sino también en la información relacionada existente.

## Fuerza de asociación

Se refiere a la magnitud con la que se incrementa el riesgo de desarrollar un efecto cuando se presenta una exposición. A mayor incremento, más fuerza de asociación y mayor seguridad de causalidad. Son tres las medidas utilizadas con más frecuencia para expresar la fuerza de una asociación: riesgo relativo (RR), riesgo atribuible (RA) y fracción etiológica (FE).

Considere la figura 11-4 (tabla de contingencia de  $2 \times 2$ ). El RR establece la razón existente entre el riesgo de los expuestos (incidencia del efecto en los expuestos) y el riesgo de los no expuestos (incidencia del efecto en los no expuestos).

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

**Cuadro 11-1.** Criterios de causalidad de Bradford-Hill

|                            |                                                                                                                                                                                                                       |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Fuerza de la asociación | Se refiere a la magnitud con la que se incrementa el riesgo de desarrollar un efecto cuando se presenta una exposición.                                                                                               |
| 2. Consistencia            | Observación repetida de una asociación en poblaciones diferentes, bajo circunstancias diferentes, y reproducida por distintos investigadores utilizando diseños distintos.                                            |
| 3. Especificidad           | Una causa conduce a un efecto único, no a efectos múltiples.                                                                                                                                                          |
| 4. Temporalidad            | La causa precede en el tiempo al efecto.                                                                                                                                                                              |
| 5. Gradiente biológico     | Presencia de una curva de dosis-respuesta. A mayor exposición mayor riesgo de enfermar.                                                                                                                               |
| 6. Plausibilidad           | Factibilidad biológica de la hipótesis.                                                                                                                                                                               |
| 7. Coherencia              | Una interpretación de causa y efecto para una asociación no debe entrar en conflicto con lo que se sabe de la historia natural y la biología de la enfermedad.                                                        |
| 8. Evidencia experimental  | Cuando existe sustento de la relación entre el factor de riesgo y la enfermedad a través de estudios como el ensayo clínico controlado. Rara vez existe evidencia en humanos.                                         |
| 9. Analogía                | Cuando es posible establecer una relación entre el factor de riesgo en estudio y otras enfermedades similares, se dice que hay equivalencia de asociación. Refuerza la credibilidad de que una asociación sea causal. |

Fuente: Tomado de Rothman K. *Epidemiología moderna*. Díaz de Santos, España, 1987.

Esta fórmula sólo es aplicable a los estudios prospectivos, ya sea experimentales (ensayos aleatorios) o de observación (de cohorte), donde es posible estimar la incidencia del efecto.

En los estudios de casos y controles se utiliza una estimación indirecta del riesgo relativo, que se conoce como razón de productos cruzados o razón de momios (RM).

$$RM = \frac{ad}{bc}$$

|            |          | Efecto                                                          |                                                                    |
|------------|----------|-----------------------------------------------------------------|--------------------------------------------------------------------|
|            |          | Presente                                                        | Ausente                                                            |
| Exposición | Presente | Número de sujetos expuestos que desarrollan el efecto<br>(a)    | Número de sujetos expuestos que no desarrollan el efecto<br>(b)    |
|            | Ausente  | (c)<br>Número de sujetos no expuestos que desarrollan el efecto | (d)<br>Número de sujetos no expuestos que no desarrollan el efecto |

**Figura 11-4.** Tabla de contingencia de 2 × 2.

Un RR o RM con valor de 1 se considera “nulo”, ya que refleja una razón entre expuestos y no expuestos de 1:1. Un RR o RM con valor de 3 o más tiene significación clínica, dado que implica un riesgo de desarrollar el efecto dos veces mayor en los sujetos expuestos que en los no expuestos. Es importante que se distinga esta significación clínica de la significación estadística, ya que la primera está dada por la magnitud del RR (o RM), y la segunda, por la probabilidad (o valor  $p$ ) de que el RR (o RM) encontrado sea real.

El RA (o diferencia de riesgos) permite distinguir el efecto absoluto de la exposición y expresa la proporción de individuos expuestos que, por efecto de la exposición, desarrollarán el efecto.

$$RA = \frac{a}{a+b} - \frac{c}{c+d}$$

La FE (o porcentaje de riesgo atribuible) permite estimar la proporción del riesgo o el efecto observado en los sujetos expuestos, que es atribuible a la exposición.

$$FE = \frac{RR-1}{RR} \text{ o } \frac{RA}{a/(a+b)}$$

## Consistencia

Una asociación es consistente si es reproducida por diferentes investigadores, utilizando distintos diseños en medios diferentes.

## Especificidad

La especificidad de una causa significa que una sola exposición genera un solo efecto y viceversa. Esto no se aplica a los factores de riesgo, ya que pueden ser varios los que contribuyan a la aparición de un solo efecto o es posible que un solo factor de riesgo sea el que contribuya a la aparición de muchos efectos.

## Temporalidad

Para que un factor causal o de riesgo se considere como tal, es indispensable que la exposición al mismo anteceda al efecto. Como se mencionó, esta relación puede demostrarse fácilmente en los estudios prospectivos, pero es posible que plantee serios problemas en estudios de casos y controles, así como en los transversales.

## Gradiente dosis-respuesta

Cuando la asociación entre una posible causa y un efecto no es muy convincente, conviene explorar el gradiente de dosis-respuesta. Se espera que a mayor magnitud o intensidad de exposición, mayor sea la magnitud o intensidad del efecto.

## Plausibilidad o credibilidad biológica

Se refiere a la concordancia de la asociación con el conocimiento biológico que se tiene en la actualidad. Este conocimiento se refiere a la comprensión de respuestas a diferentes niveles, de la célula a los organismos complejos. En este caso, es importante tener en cuenta que los resultados obtenidos en animales de experimentación no son directamente extrapolables a los seres humanos.

## Coherencia

La interpretación de causa y efecto de la asociación no debe contradecir lo que se acepta sobre la historia natural y la biología de la enfermedad.

## Evidencia experimental

Cuando existe sustento de la relación entre el factor de riesgo y la enfermedad a través de estudios experimentales (por lo general realizados en animales de experimentación).

## Analogía

Es la posibilidad de establecer una relación entre el factor de riesgo en estudio y otras enfermedades similares; se dice que hay equivalencia de asociación. Refuerza la credibilidad de que una asociación sea causal. Algunos autores incluyen, además, el siguiente punto:

## Credibilidad epidemiológica

Si existe asociación, cabe esperar que la distribución geográfica de la exposición sea paralela a la distribución del efecto.

Con respecto al trabajo de Martin y Bracken, el diseño que se utilizó fue un estudio de cohortes. Son dos los tipos de sesgo que pudieron afectar el estudio; específicamente, los sesgos de selección y medición.

Así, de 6219 candidatas para el estudio, sólo se incluyeron 3 858 (62%). Si de las exclusiones/eliminaciones se justifican 208 por ausencia de embarazo al momento del reclutamiento, 116 por abortos u óbitos y 44 por embarazos múltiples, el porcentaje de exclusión/eliminación se reduce de 38 a 34%. Con este porcentaje tan alto (el máximo aceptable es de 20%), cabe preguntar lo siguiente: ¿cuán representativa fue la muestra estudiada de la población que intentaba evaluarse? ¿El 34% perdido podía haber brindado información que proporcionara resultados diferentes a los obtenidos? En cuanto al sesgo de medición, el estudio no consideró un seguimiento que garantizara la estabilidad del estado de exposición. De esta manera, mujeres inicialmente clasificadas como expuestas pudieron haber vuelto negativa su exposición durante los meses de gestación y las que eran negativas quizá se tornaron positivas. El efecto de estos resultados falsos positivos y negativos pudo llevar a subestimar la magnitud de la asociación.

Si se considera que de las 853 mujeres 2.34% tuvo producto de bajo peso al nacer, entonces hubo 20 productos de bajo peso en este grupo, ya que  $(853 \times 2.34)/100 = 19.9$ .

De las 1 620 mujeres que no estuvieron expuestas, 0.86% de ellas parió productos de bajo peso al nacer; de tal modo, hubo 14 productos de bajo peso en este grupo, puesto que  $(1\,620 \times 0.86)/100 = 13.9$ . Con esta información se construyó una tabla de contingencia de  $2 \times 2$  (figura 11-5) para calcular en qué proporción el riesgo de tener un niño de bajo peso al nacer fue mayor en una mujer embarazada expuesta de manera pasiva al humo de tabaco con respecto al riesgo de una mujer embarazada no expuesta a este factor.

Al aplicar la prueba de chi (ji) cuadrada, ésta proporciona un valor de 7.95 que a su vez corresponde a un valor  $p < 0.01$ . Dado que la asociación entre tabaquismo pasivo y bajo peso al nacer es estadísticamente significativa, puede concluirse que ésta es real.

Si se considera que al ajustar el RR a los factores de confusión éste disminuye a 2.17, ¿esto puede tener significado clínico?

El RR de 2.71 o 2.17 del estudio es real, pero no clínicamente significativo. Sin embargo, al calcular los IC 95% de estos mismos RR, aquéllos varían de un mínimo de 1.05 a un máximo de 5.36, lo cual quiere decir que los riesgos obtenidos pueden ser tan bajos como 1.05 o tan altos como 5.36. Ante esta perspectiva, no es posible descartar la exposición materna pasiva al humo de tabaco como factor de riesgo de importancia clínica para el desarrollo de un producto de bajo peso.

El RA del estudio es de  $(20/853) - (14/1\,620) = 0.015$  y refleja que de cada 100 niños nacidos de madres fumadoras pasivas, el bajo peso al nacer de dos productos es atribuible al tabaquismo pasivo de la madre. Estos dos productos representan, a su vez, 63% de todos aquellos con bajo peso al nacer paridos por madres expuestas al humo del cigarro  $[FE = (2.713 - 1)/2.713 = 0.631]$ .

**Producto de bajo peso al nacer**

|                                     |    | Sí    | No    |       |
|-------------------------------------|----|-------|-------|-------|
| Exposición pasiva al humo de tabaco | Sí | 20    | 833   | 853   |
|                                     | No | 14    | 1 606 | 1 620 |
|                                     | 34 | 2 439 | 2 473 |       |

$$RR = \frac{20/(20 + 833)}{14/(14 + 1\,606)} = \frac{20/853}{14/1\,620} = 2.713$$

**Figura 11-5.** Tabla de contingencia de  $2 \times 2$  para calcular el riesgo de dar a luz un niño de bajo peso al nacer en mujeres embarazadas expuestas al humo de tabaco. RR = riesgo relativo.

Por otro lado, si se considera qué tan bajos fueron los pesos de los productos de bajo peso al nacer en el grupo expuesto, se encuentra que el déficit promedio fue de 30 g, cuya significación clínica es discutible. Existe en la literatura mundial una serie de trabajos que confirma la asociación de tabaquismo materno y bajo peso del producto. La variación en el trabajo de Martin y Bracken es la exposición pasiva al humo de tabaco, la cual, a diferencia de la exposición activa, ofrece mayor variación en términos de intensidad.

La relación temporal de exposición y efecto no está muy clara en el estudio, dado que las candidatas eran captadas cuando tenían tiempo variable de gestación, durante el cual el estado de exposición pudo ser diferente al registrado.

No se proporcionan resultados de niveles distintos de exposición. Sin embargo, está documentado en la literatura existente que el peso del producto al nacer es más bajo a medida que aumenta el número de cigarrillos consumidos por la madre.

En la mujer gestante se ha observado que el tabaquismo aumenta las concentraciones de carboxihemoglobina, disminuye la liberación de oxígeno sanguíneo y produce vasoconstricción, todo lo cual ocasiona hipoxia fetal lo cual, a su vez, causa retardo de crecimiento.

Son varios los factores cuya influencia sobre el producto se ha postulado, como multiparidad, medidas anticonceptivas, enfermedades sistémicas concomitantes, etcétera. Ello permite entender que la asociación entre tabaquismo materno y peso del producto no es específica.

Por otro lado, en el estudio de metaanálisis realizado por Salmasi, se halló que el peso de los recién nacidos de madres expuestas fue menor (diferencia de medias de 60 g, 95% intervalo de confianza (IC)  $-80$  a  $-39$  g), con una tendencia hacia el aumento de bajo peso al nacer (BPN,  $<2\,500$  g; RR 1.16; 95% IC 0.99-1.36); la duración de la gestación y el nacimiento pretérmino fue similar (diferencia de medias 0.02 semanas, 95% IC  $-0.09$  a 0.12 semanas y RR 1.07; 95% IC 0.93-1.22). En los recién nacidos expuestos aumentó el riesgo de anomalías congénitas (OR 1.17; 95% IC 1.03-1.34) y se observó una tendencia a tener menor circunferencia craneana ( $-0.11$  cm; 95% IC  $-0.22$  a 0.01 cm). Conclusiones: las mujeres expuestas al tabaquismo pasivo mostraron mayor riesgo de tener hijos con bajo peso al nacer, anomalías congénitas, mayor longitud y tendencia a tener menor circunferencia craneal.

---

## Bibliografía

- Department of Clinical Epidemiology and Biostatistics. McMaster University: How to read clinical journals: IV. To determine etiology or causation. *Can Med Assoc J* 124:985-990. 1981.
- Feinstein AR. Clinical Biostatistics-XLVII. Scientific standards *vs.* statistical associations and biologic logic in the analysis of causation. *Clin Pharmacol Ther* 25(4):481-492. 1979.
- Feinstein AR. Clinical Biostatistics-XLVIII. Efficacy of different research structures in preventing bias in the analysis of causation. *Clin Pharmacol Ther* 26(1):129-141. 1979.
- Kimberly Yolton, Yingying Xu, Jane Khoury, Paul Succop, Bruce Lanphear, Dean W. Associations between second hand smoke exposure and sleep patterns in children. *Pediatrics* 125:e261-e268. 2010.



- Hennekens ChC, Buring JE. Epidemiology in medicine. Statistical association and cause-effect relationships. Little, Brown and Company, Boston, EUA, pp. 30-53. 1987.
- Martin TR, Bracken MB. Association of low birth weight with passive smoke exposure in pregnancy. *Am J Epidemiol* 124(4):633-642. 1986.
- Mausner JS, Bahn AK. Epidemiologic concepts and models. En: *Epidemiology, an introductory text*. WB Saunders, Boston, EUA, pp. 21-42. 1987.
- Mausner JS, Bahn AK. The search for causal relations: observational studies. En: *Epidemiology, an introductory text*. W/B Saunders, Boston, EUA, 91-111. 1987.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*, 3a. ed. Lippincott Williams and Wilkins, Boston, EUA. 2008.
- Sacket D, Haynes B, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Little Brown, Boston, EUA, pp. 3-158. 1985.
- Salmasi G, Grady R, Jones J, McDonald SD, Knowledge Synthesis Group. Environmental tobacco smoke exposure and perinatal outcomes: a systematic review and meta-analyses. *Acta Obstet Gynecol Scand* 89(4):423-441. 2010.
- Siegel S. *Estadística no paramétrica. El caso de dos muestras independientes*. Trillas, México, pp. 120-137. 1972.
- Stehbens WE. The concepts of cause to disease. *J Chronic Dis* 38(11):947-950. 1985.

# Capítulo 12

## Diagnóstico y evaluación de pruebas diagnósticas

*Laura Moreno Altamirano*



### Introducción

El diagnóstico puede considerarse como el más importante resultado de la práctica médica, ya que su establecimiento conduce al tratamiento y al pronóstico del padecimiento, mismo que resulta un problema complejo en ese ejercicio de la clínica.

El diagnóstico se puede definir como una hipótesis acerca de la naturaleza de la enfermedad de un paciente, que se deriva de observaciones a través del uso de la inferencia. Varios autores reconocen que la presencia de una enfermedad en un individuo a menudo no puede determinarse con certeza; agregan, además, que el objetivo del médico no es alcanzar la certeza sino reducir el nivel de incertidumbre lo suficiente como para tomar la decisión terapéutica.

El proceso general que da lugar al diagnóstico se realiza en dos etapas; en la primera se establece una presuposición, sospecha o hipótesis de existencia de la enfermedad. La segunda se dirige al seguimiento de la sospecha clínica y a verificar si ésta corresponde a la realidad.

En este proceso, de enorme complejidad, existe un gran número de fuentes de incertidumbre que transitan por una amplia gama de cuestiones, como son: que el conjunto de síntomas y signos en un paciente puede ser compatible con más de una enfermedad, que existen variaciones biológicas a veces importantes entre un enfermo y otro, que los instrumentos suelen ser imprecisos, factores inherentes a la experiencia, formación y la acuciosidad del médico, y que los pacientes son inexactos para recordar sucesos pasados, entre otros.

---

Agradecimientos a la Dra. Rosario Velasco Lavín por su apoyo en la revisión del caso clínico. Y al Ing. David Limón por su apoyo técnico y por el diseño de las figuras.

Si bien la historia clínica dentro del acto médico es un elemento insustituible y la base fundamental para instituir o descartar el diagnóstico de las enfermedades y constituye el sustento principal para establecer su pronóstico y tratamiento, es innegable que las pruebas de laboratorio y gabinete deben considerarse como excelentes recursos de apoyo para el médico.

No obstante, es importante tener en mente que, si bien las diferentes fases por las que pasa la relación entre médico y paciente, según el momento de la atención, requieren de distintos matices durante el acto médico, el paciente acude al médico solicitando remedio para su padecer, y muchas veces el interés del médico se centra en el examen científico del proceso patológico, por lo que la interacción y el vínculo pasan a un segundo plano, estableciendo una relación con el paciente más como un “objeto de estudio” que como un ser humano, lo cual distorsiona el verdadero sentido de esta relación.

### Caso clínico

---

En la consulta externa de un servicio de gastroenterología se atendió a un paciente de 55 años, quien refirió presentar desde seis meses antes: disminución del apetito, pérdida de peso, fatiga, debilidad, inflamación del abdomen y sensación de llenura. Además señaló que en las últimas semanas había presentado episodios de fiebre y dolor en la parte superior del lado derecho del abdomen que se extendía a la espalda y al hombro. Entre sus antecedentes de importancia el paciente refiere que a los 30 años se le diagnosticó hepatitis B sin control. Decidió acudir al servicio médico debido a que la piel y lo blanco de los ojos se le pusieron amarillos y la orina muy oscura.

Se le indicaron pruebas de función hepática, serología para hepatitis B y niveles de alfa fetoproteína. Las pruebas hepáticas resultaron alteradas, con elevación de aminotransferasas, fosfatasa alcalina, bilirrubinas y deshidrogenasa láctica; se corrobora hepatitis B crónica por serología y niveles elevados de alfa fetoproteína.

Con esta información, ¿qué diagnósticos probables podría establecer?; ¿qué otras pruebas realizaría para confirmarlos o descartarlos?

---

Ante la evidencia clínica o sospecha fundamentada de la existencia de una enfermedad, es decir, cuando por medio del interrogatorio clínico y la exploración física se plantea un posible diagnóstico, resulta necesario confirmarlo o descartarlo. Es un hecho indiscutible que antes de indicar procedimientos diagnósticos debe existir cierta probabilidad de que el paciente tenga la enfermedad. En otras palabras, los procedimientos diagnósticos deben sustentarse firmemente en la información obtenida a través del acto clínico.

Las pruebas diagnósticas además son de gran utilidad para la detección de casos, para establecer el pronóstico del padecimiento o bien para medir o determinar el efecto de un tratamiento. Su interacción determina la base del éxito o fracaso de la acción médica. Los procedimientos de diagnóstico se han ido desarrollando cada vez más, quizás en tecnificación extrema; en la actualidad existe una amplia gama de recursos diagnósticos que obligan al médico clínico a saber cuándo solicitarlos, cuál o cuáles de ellos son realmente necesarios, con qué secuencia decidir su uso y sobre todo con qué seguridad aceptarlos como válidos en una adecuada interpretación.

Es decir, el desarrollo tecnológico, los equipos nuevos y los recientes procedimientos de diagnóstico de laboratorio y gabinete han abierto un campo sumamente amplio en el área de la investigación, lo que lleva implícita la necesidad de utilizarlos en la práctica diaria, buscando el mayor beneficio tanto para los pacientes como para las instituciones médicas.

Se llama **prueba diagnóstica** (PD) a cualquier proceso, más o menos complejo, que pretenda determinar en un paciente la presencia de cierta condición, supuestamente patológica, no susceptible de ser observada directamente (con alguno de los cinco sentidos elementales).

Los procedimientos diagnósticos han sido aceptados como útiles, necesarios y en la mayor parte de las ocasiones indispensables; de ahí el creciente interés por analizar o evaluar no sólo su utilidad sino también su **validez o eficacia**.

#### **Validez**

A la validez también se le llama certeza, conformidad o precisión. Se refiere a qué tanto concuerda el resultado de la medición con el valor real. Cuando existe un patrón o estándar de oro, se refiere a qué tanto concuerda ese estándar de oro con el resultado de la medición (p. ej., si concuerda el resultado de una prueba diagnóstica con el de una biopsia o un cultivo). El concepto de "validez" es más amplio; también se refiere a qué tanto refleja la verdad esa medición (capítulo 9).

#### **Eficacia**

Definida desde el punto de vista estadístico como la escisión con la que una prueba diagnóstica estima el parámetro de interés.

Desde la perspectiva de la Epidemiología Clínica se han propuesto alternativas metodológicas que permiten conocer con mayor objetividad la eficacia de las pruebas diagnósticas; la finalidad de obtener este conocimiento es contar con elementos para tomar decisiones clínicas de manera más fundamentada y aplicables a la práctica médica cotidiana. En este capítulo se enfatizará sobre la importancia de validar las pruebas diagnósticas y su utilidad práctica.

Existen diferentes tipos de procedimientos diagnósticos que se utilizan en forma general y otros que corresponden a la práctica especializada. De acuerdo con su eficacia, se pueden agrupar como se muestra en la figura 12-1.

Las pruebas no factibles ni pertinentes carecen de importancia clínica, ya que, o no se dispone de ellas o resultan peligrosas para los pacientes.

Lo ideal es contar con procedimientos de alta eficacia, factibles (esto es, la disponibilidad de los recursos humanos, físicos, técnicos, materiales, etcétera, para practicar la prueba diagnóstica) y pertinentes (que consiste en la seguridad en cuanto al riesgo para la salud

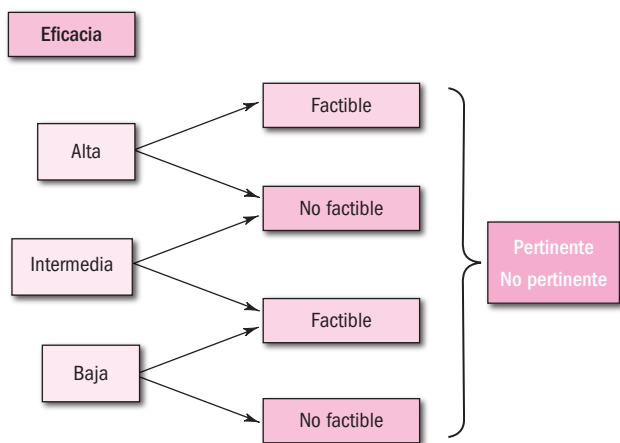


**Figura 12-1.** Agrupación de las pruebas diagnósticas de acuerdo a su eficacia y pertinencia.

del paciente en relación con el beneficio que se le ofrece), definidas con estas características por medio de la experiencia documentada y la experimentación fundamentada. Como esta situación no siempre es posible, cuando se cuenta con una prueba eficaz, factible, pero no pertinente es necesario buscar otra alternativa y evaluar el grado de eficacia en relación con otra que recibe el nombre de estándar diagnóstico ideal, prueba de fuego o estándar de oro.

El **estándar diagnóstico ideal** es el procedimiento o criterio utilizado para diagnosticar con certeza una enfermedad y también como referencia para evaluar otras pruebas. No es posible tener en todos los casos un estándar ideal, es decir, una prueba que diagnostique correctamente al 100% de los individuos. Entonces se selecciona la prueba más eficaz y pertinente con la que se cuente y se compara con ella el método diagnóstico en cuestión.

La biopsia, por ejemplo, es considerada en general como uno de los procedimientos diagnósticos más confiables; sin embargo, no siempre es factible o pertinente realizarla, situación en la que se busca otra alternativa, como podrían ser los estudios de laboratorio o de imagen, o bien la evolución clínica del paciente.

Se presentan diversas circunstancias en que las pruebas diagnósticas tienen restricciones y se hace necesario seleccionar la más adecuada; se citan las siguientes condiciones en las que la evaluación es insoslayable:

- Cuando el estado del paciente no permite el uso de la prueba más eficaz por el riesgo inherente al estudio y sus complicaciones potenciales.
- Cuando se cuenta con una prueba “nueva” que puede ser una alternativa menos riesgosa, pero con la que se carece de la experiencia necesaria.
- Cuando se cuenta con una prueba más barata o de operación más sencilla.

(continúa)

(Continuación)

- En el caso que existan varios procedimientos, para elegir el mejor.
- Cuando no es factible contar con las pruebas más confiables y se desea conocer el grado de eficacia de la existente.
- Para establecer la secuencia de varias pruebas confiables.

Las pruebas diagnósticas disponibles en la actualidad tienen diferente grado de eficacia entre ellas mismas y para las diferentes enfermedades. Por ejemplo, la eficacia de la prueba de esfuerzo es diferente que la de la coronariografía para establecer el diagnóstico de infarto al miocardio. Asimismo, la prueba de esfuerzo tiene diferente eficacia para establecer el diagnóstico de hipertensión arterial e isquemia miocárdica. Por otro lado, los procedimientos diagnósticos disponibles para un mismo padecimiento pueden ser múltiples o muy limitados.

La evaluación de la eficacia de una prueba diagnóstica comienza por la cuantificación o estimación de la magnitud de los errores que pueden cometerse o, su inverso, la magnitud de los aciertos que se realizan al intentar establecer un diagnóstico a partir de los resultados que brinde dicho procedimiento. Para ello habría que conocer su **sensibilidad y especificidad**.

En 1947, Yerushalmy introdujo los términos de **sensibilidad** y **especificidad** como indicadores estadísticos que evalúan el grado de eficacia inherente a una prueba diagnóstica. La sensibilidad y la especificidad son las medidas tradicionales y básicas del valor diagnóstico de una prueba. Miden la discriminación diagnóstica de una prueba en relación con un criterio de referencia, que se considera la verdad. Estos indicadores en principio permiten comparar directamente la eficacia de una prueba con la de otras y esperar resultados similares cuando son aplicadas en diferentes países, regiones o ámbitos.

El clínico debe conocer con cierta precisión tales valores en las pruebas que utiliza de manera cotidiana para darles el peso adecuado en sus decisiones. Por otra parte, el médico que realiza una investigación sobre una prueba diagnóstica debe determinar su sensibilidad y su especificidad; de otra manera, su trabajo se considerará irrelevante.

Existe una amplia gama de posibles investigaciones que se pueden realizar con las pruebas diagnósticas; algunas de ellas son las siguientes:

- Determinar la eficacia de la prueba (sensibilidad, especificidad y valores de predicción positivo y negativo) para diagnosticar diversos padecimientos.

(continúa)

(Continuación)

- Valorar si la prueba es útil cuando es aplicada a grupos numerosos, en estudios de tamiz o escrutinio.
- Definir límites de normalidad a través de la prueba y los puntos de corte para el diagnóstico de una enfermedad.
- Comparar los resultados de varias pruebas para conocer cuál es la mejor para identificar una enfermedad o para descartarla.
- Conocer las alteraciones cualitativas que se encuentran en las pruebas en diversas enfermedades.
- Definir el orden en que deben practicarse los estudios de diagnóstico en los pacientes con determinado padecimiento.

## Eficacia de una prueba

Para que una prueba se considere eficaz es indispensable que tanto la especificidad como la sensibilidad se acerquen al 100 por ciento.

La **sensibilidad** está dada por la proporción de individuos en los que la prueba resulta positiva cuando realmente tienen la enfermedad.

La **especificidad** está dada por la proporción de individuos en los que la prueba resulta negativa en ausencia de la enfermedad.

No es correcto afirmar que una prueba es eficaz cuando es muy sensible; es decir, cuando en casi todos los enfermos en quienes se aplica, el resultado es positivo.

Para que una prueba se considere adecuada, se requiere además que sea específica; esto es, que también pueda detectar a los sujetos no enfermos mediante resultados negativos.

Vale la pena destacar que tanto la sensibilidad como la especificidad se refieren a una enfermedad determinada. Una prueba puede tener diferentes valores en relación con distintas entidades. Esto es particularmente cierto en las pruebas de diagnóstico que utilizan imágenes: radiografías, gammagrafías, estudios de ultrasonido y tomografía axial, y otros.

## Cálculo de la sensibilidad y la especificidad

Para calcular estos valores es necesario comparar a un grupo de individuos con la enfermedad con un grupo de individuos sin ella.

La definición de **enfermedad** se debe basar en el cuadro clínico y la prueba o pruebas consideradas estándar ideal, que se aplican tanto al grupo de enfermos como al de

**Cuadro 12-1.** Normas generales para la aplicación de los programas de detección. Criterios para seleccionar una prueba de detección

| Resultados de la prueba | Enfermedad o estándar ideal |                      |
|-------------------------|-----------------------------|----------------------|
|                         | Presente                    | Ausente              |
| +                       | Verdaderos positivos        | Falsos positivos     |
| –                       | Falsos negativos            | Verdaderos negativos |

no enfermos. Es necesario, además, especificar la gravedad o la etapa en que se encuentra la enfermedad para que la extrapolación de resultados sea adecuada.

Los resultados obtenidos de acuerdo con el estándar ideal o diagnóstico de certeza son: positivo, cuando se considera que el individuo tiene la enfermedad; negativo, cuando se comprueba que no la presenta. La interrelación entre estos resultados y los de la prueba en estudio se expresa en el cuadro 12-1.

Existen cuatro resultados: cuando en presencia de enfermedad la prueba es positiva (verdadero positivo), y cuando en ausencia de la enfermedad, la prueba es negativa (verdadero negativo). Por otro lado, cuando en ausencia de enfermedad la prueba es positiva (falso positivo), y cuando en presencia del padecimiento, ésta es negativa (falso negativo).

Después, dichos resultados permitirán elaborar una tabla de  $2 \times 2$  o tabla de contingencia.

|                   |   | Enfermedad o estándar ideal |       |               |
|-------------------|---|-----------------------------|-------|---------------|
|                   |   | +                           | –     |               |
| Prueba de estudio | + | a                           | b     | a + b         |
|                   | – | c                           | d     | c + d         |
|                   |   | a + c                       | b + d | a + b + c + d |

Sensibilidad =  $a/(a + c)$  =

Especificidad =  $d/(b + d)$  =

a = número de casos verdaderos positivos

b = número de casos falsos positivos

c = número de casos falsos negativos

d = número de casos verdaderos negativos

a + c = total de casos con enfermedades independientemente de los resultados de la prueba diagnóstica

b + d = total de casos sin la enfermedad independientemente de los resultados de la prueba diagnóstica

a + b = total de casos positivos a la prueba independientemente de tener o no la enfermedad

c + d = total de casos negativos a la prueba independientemente de tener o no la enfermedad

a + c + b + d = total de casos estudiados.



$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{\text{Total de casos con la enfermedad}} = \frac{a}{a + c}$$

**Sensibilidad.** En el numerador aparecen los enfermos en quienes la prueba fue positiva (verdaderos positivos), y en el denominador, el número total de sujetos con la enfermedad, independientemente de que la prueba haya sido positiva (verdaderos positivos) o negativa (falsos negativos).

$$\text{Especificidad} = \frac{\text{Verdaderos negativos}}{\text{Total de casos sin la enfermedad}} = \frac{d}{b + d}$$

**Especificidad.** En el numerador se coloca el número de casos negativos entre el total de individuos que no tenían la enfermedad (verdaderos negativos), y en el denominador el total de personas sin la enfermedad con y sin la prueba positiva (falsos positivos y verdaderos negativos, respectivamente).

La sensibilidad y la especificidad se presentan como porcentaje; a medida que ambos índices se acercan al 100%, se considera que la prueba es más eficaz. Es decir, una prueba diagnóstica que tiene alta sensibilidad (96%) y baja especificidad (41%) no se considera adecuada, ya que aunque detecte 96% de enfermos, su capacidad para identificar a los sanos es limitada.

Las pruebas muy sensibles son útiles para descartar padecimientos y si son accesibles e inocuas se eligen en las fases iniciales del estudio de los pacientes.

Asimismo, una prueba con especificidad de 96% y sensibilidad de 41% es útil para detectar 96% de los sanos, pero sólo identifica 41% de los enfermos.

Vale la pena enfatizar que cuando la enfermedad que se sospecha es de mal pronóstico pero tratable y, por tanto, es muy importante hacer el diagnóstico, se elige una prueba con alta sensibilidad (con un porcentaje muy bajo de resultados falsos negativos).

Por otro lado, en el caso de un paciente con una neoplasia maligna se requiere una prueba muy específica que tenga pocos resultados falsos positivos, para que no origine problemas al individuo o a su familia sin ninguna justificación. Si ya se ha establecido un diagnóstico probable en un paciente y se quiere confirmar, es preferible elegir la prueba que tenga el índice de especificidad más alto.

Con el propósito de esclarecer estos conceptos, retome el caso clínico inicial. Se estableció que el paciente tenía una hepatopatía crónica, y se debía descartar la presencia de hepatitis crónica viral o autoinmune, cirrosis o carcinoma hepático. Dadas las alteraciones y el antecedente de hepatitis B, se sospecha de carcinoma hepatocelular, para lo cual y, de acuerdo con las recomendaciones de la Asociación Norteamericana para el Estudio de las Enfermedades del Hígado (*American Association for the Study of Liver Disease*), se plantea la realización de estudios de imagen como tomografía axial computarizada y resonancia magnética y/o biopsia hepática percutánea, la que se considera como estándar de oro.

Para ello se revisó un estudio multicéntrico en el que se identificaron 60 pacientes con hepatopatía crónica, candidatos a trasplante hepático, en quienes se deseaba descartar la presencia de carcinoma hepático.

Antes de someterlos a cirugía se les practicó una serie de estudios para confirmar el diagnóstico de carcinoma hepatocelular (CHC).

Los criterios diagnósticos para CHC asociados a cirrosis fueron establecidos desde 1954 por Edmonson y colaboradores. No obstante, el desarrollo de técnicas de imagen no invasivas en sí mismas y como auxiliares para la obtención de biopsias por aspiración dirigidas, además de exámenes inmunoquímicos como la determinación de alfa fetoproteína (AFP), abren una serie de posibilidades diagnósticas muy pertinentes.

A estos pacientes, en primer lugar se les indicó una determinación de alfa fetoproteína y además a todos se les practicó una biopsia por aspiración, considerada para este estudio como “prueba estándar ideal”.

Con la finalidad de conocer la sensibilidad y especificidad de esta prueba, se elaboró la tabla que sigue, donde se presentan los resultados obtenidos:

|                          |   | <b>Biopsia por aspiración</b> |      |    |
|--------------------------|---|-------------------------------|------|----|
|                          |   | +                             | -    |    |
| <b>Alfa fetoproteína</b> | + | 19a                           | b 20 | 39 |
|                          | - | 2c                            | d 19 | 21 |
|                          |   | 21                            | 39   | 60 |

$$\text{Sensibilidad} = 19/21 \times 100 = 90\%$$

$$\text{Especificidad} = 19/39 \times 100 = 48\%$$

Se encontró que la alfa fetoproteína es una prueba muy sensible (90%), pero poco específica (48%). Sin embargo, niveles altos de alfa fetoproteína pueden encontrarse además en otros padecimientos como hepatitis, regeneración hepatocelular activa, tumores germinales, entre otros, es decir, tiende a sobrediagnosticarse. La especificidad de 48% es un valor considerado bajo, ya que no identifica claramente a quienes no tienen la enfermedad. Este tipo de pruebas es de gran utilidad para la detección oportuna de enfermedades en población de alto riesgo.

Por tanto, cabe concluir que es eficaz para identificar a los que tienen cáncer hepático, pero no para descartar a los que no lo tienen.

Luego, a este grupo de pacientes se les realizó estudio de ultrasonido hepático, cuyos resultados se muestran en el siguiente cuadro:

|                    |   | <b>Biopsia por aspiración</b> |      |    |
|--------------------|---|-------------------------------|------|----|
|                    |   | +                             | -    |    |
| <b>Ultrasonido</b> | + | 16a                           | b8   | 24 |
|                    | - | 5c                            | d 31 | 36 |
|                    |   | 21                            | 39   | 60 |

$$\text{Sensibilidad} = 16/21 \times 100 = 76\%$$

$$\text{Especificidad} = 31/39 \times 100 = 79\%$$

Se tomó como prueba estándar ideal también la biopsia por aspiración y se observó que es una prueba menos sensible (76%) que la alfa fetoproteína, pero mucho más específica (79%). Aunque aparentemente la alfa fetoproteína es mejor en cuanto a sensibilidad, si se analizan los datos de los cuadros anteriores, en la casilla b se observa que el ultrasonido sólo sobrediagnosticó a ocho individuos, a diferencia de la alfa fetoproteína que sobrediagnosticó a 20, es decir, 29 y 51%, respectivamente. Es aquí donde el criterio médico debe prevalecer para la adecuada interpretación sobre la utilidad de una prueba, sin dejarse llevar únicamente por los datos numéricos.

Más tarde se llevó a cabo el trasplante hepático, de donde se obtuvieron lesiones nodulares que se analizaron histopatológicamente. Los hallazgos se compararon con los de la biopsia por aspiración, tomando como estándar ideal el estudio histopatológico de espécimen obtenido durante la intervención quirúrgica o bien en estudio *post mortem*:

|                        |   | Histopatológica del espécimen |    |    |
|------------------------|---|-------------------------------|----|----|
|                        |   | +                             | −  |    |
| Biopsia por aspiración | + | 17                            | 4  | 21 |
|                        | − | 2                             | 37 | 39 |
|                        |   | 19                            | 41 | 60 |

$$\text{Sensibilidad} = 17/19 \times 100 = 89\%$$

$$\text{Especificidad} = 37/41 \times 100 = 90\%$$

Se encontró que la biopsia por aspiración es muy sensible y específica, 89 y 90%, respectivamente. Sin embargo, no llega al 100%, debido a la influencia de una serie de factores, en especial los criterios diagnósticos y la nomenclatura heterogénea que se utilizó para describir las lesiones histológicas.

La posible explicación podría ser que los criterios establecidos por Edmonson no habían sido revisados; en la actualidad se han propuesto otras nomenclaturas, como la de Ferrell y colaboradores, que las agrupa en: *a*) nódulo de regeneración, *b*) lesiones limítrofes y *c*) adenoma hepatocelular bien diferenciado. O la de la OMS, la de Barcelona, o bien la más recientemente propuesta por la *American Association for the Study of Liver Disease* (AASLD), basada en el tamaño de los nódulos hepáticos.

**Carcinoma hepatocelular.** Grados: la gradación de la OMS equivalente a la de Edmonson y Steiner

- Bien diferenciado: el grado I de Edmonson es el observado en el CHC pequeño o precoz.
- Moderadamente diferenciado.
- Pobremente diferenciado.
- Indiferenciado.

Para conocer con más profundidad las razones de una impresión diagnóstica diferente es necesario indagar sobre los criterios utilizados, así como en la concordancia existente entre los observadores, aspectos que se revisan en el capítulo 9.

## Confiabilidad de los valores de sensibilidad y especificidad

Después de obtener la sensibilidad y especificidad de una prueba, el médico se pregunta: ¿cuánto puedo confiar en estos valores? y ¿qué tan adecuado es extrapolar estos resultados?

Para contestar esas preguntas es necesario asegurarse de que se tomaron en cuenta los siguientes aspectos:

- Definición clara y estricta de los criterios de inclusión de los pacientes.
- Cálculo adecuado del tamaño de muestra; se recomienda obtenerlo a partir de las estrategias señaladas en el capítulo 17. O mediante tablas elaboradas por García Romero, para este fin (capítulo 12, anexo 1).
- Establecimiento de los límites de normalidad de las pruebas estudiadas, con base en valores convencionales u obtenidos por medio de **curva ROC**.
- Cálculo de la confiabilidad de los valores de sensibilidad y especificidad obtenidos.

A continuación se explica brevemente el procedimiento para calcular la confiabilidad de los valores de sensibilidad y especificidad.

Cuando se señala que la sensibilidad y especificidad de la prueba de ultrasonido en una muestra de individuos con carcinoma hepatocelular es de 76 y 79%, respectivamente, se espera que estos valores sean muy parecidos a los valores reales, que se obtendrían si se estudiara todo el universo de pacientes con este padecimiento.

A fin de asegurar con 95% de probabilidad que los datos obtenidos en el estudio son semejantes a los reales, se establecen los límites entre los cuales puede estar ese valor. Para ello se le suma y se le resta 1.96 por la desviación estándar.

La desviación estándar (DE) se calcula usando la siguiente fórmula:

$$DE = \sqrt{p \times q/n}$$

Para calcular la desviación estándar de la sensibilidad

DE Desviación estándar

p Sensibilidad  $a/(a+c)$

q Proporción de falsos positivos entre el total de no enfermos  $c/(a+c)$

n Total de individuos estudiados

La desviación estándar de la sensibilidad del ultrasonido es:

$$DE = \sqrt{\frac{0.76 \times 0.24}{60}} = 0.055$$

Para calcular la desviación estándar de la especificidad

DE Desviación estándar

- p Especificidad  $d/(b+d)$   
 q Proporción de falsos positivos entre el total de individuos enfermos  $b/(b+d)$   
 n Total de individuos estudiados

La desviación estándar de la especificidad del ultrasonido es:

$$DE = \sqrt{\frac{0.79 \times 0.18}{60}} = 0.048$$

Una vez que se conoce la desviación estándar, se pueden obtener los límites mencionados. Los límites de confianza para la sensibilidad.

$$0.76 + (1.96 \times 0.05) \text{ y } 0.76 - (1.96 \times 0.05) = \\ 0.858 \text{ y } 0.662$$

Es decir, los límites de confianza para la sensibilidad del ultrasonido están entre 85.8 y 66.2 por ciento.

Límites de confianza para la especificidad

$$0.79 + (1.96 \times 0.048) \text{ y } 0.79 - (1.96 \times 0.048) = \\ 0.884 \text{ y } 0.696$$

Así, los límites de especificidad para el ultrasonido están entre 88.4 y 69.6 por ciento.

Mientras mayor sea la muestra de donde se obtengan los datos de sensibilidad y especificidad, menores serán tanto la desviación estándar, como el intervalo de confiabilidad de los índices. Es decir, el valor de la muestra se alejará menos del valor real. El valor de la desviación estándar indica qué tanto se puede confiar en los índices que se obtienen como representativos de los valores reales.

Un aspecto complejo es el de los casos inciertos o dudosos y el de los casos no comprobados. La mayoría de los investigadores prefiere excluirllos; no obstante, García Romero y García Barrios proponen el manejo que se expone en el anexo 2 de este capítulo.

## Valores predictivos

Tan importante como conocer la sensibilidad y la especificidad, es de interés para el clínico o investigador saber cuándo la prueba es positiva en un individuo. ¿Cuál es la probabilidad de que éste realmente tenga el padecimiento?, lo que se conoce como **valor de predicción de una prueba positiva** (VP+). Asimismo, cuando la prueba es negativa, ¿cuál es la probabilidad de que el individuo no tenga la enfermedad?, conocido como valor de **predicción de una prueba negativa** (VP-).

El valor de predicción de una prueba varía en relación con la prevalencia de la enfermedad en la población estudiada, y está determinado por la sensibilidad y especificidad de la prueba. La prevalencia es la proporción de sujetos con la enfermedad en una población determinada y en un momento dado, de modo que los valores de predicción deben aplicarse a la misma población en donde se estimó la prevalencia.

Para calcular estos valores existen dos procedimientos; uno de ellos se realiza utilizando el teorema de Bayes, el que a continuación se desarrolla.

### Valor de predicción positivo

$$VP+ = \frac{S \times P}{(S \times P) + (1-E) \times (1-P)}$$

S = sensibilidad  
P = prevalencia  
E = especificidad

### Valor de predicción negativo

$$VP- = \frac{E \times (1-P)}{(E) \times (1-P) + (1-S) \times P}$$

1 - S = diferencia de la sensibilidad  
1 - P = diferencia de la prevalencia  
1 - E = diferencia de la especificidad

En el ejemplo ya mencionado se observó una sensibilidad de 0.89 (89%) y una especificidad de 0.90 (90%) para la biopsia por aspiración, valores que se emplearon en el cálculo de VP+ y VP- de esta prueba en individuos con hepatopatía crónica, y con una prevalencia para CHC de 0.20 (es decir, 20% del total de enfermos con hepatopatía pueden desarrollar CHC).

A dos de los pacientes del grupo bajo estudio, uno de 63 y otro de 61 años, ambos con cirrosis hepática de 2 y 3 años de evolución, respectivamente, se les practicó una biopsia por aspiración. En el primer paciente resultó compatible con CHC; es decir, la prueba fue positiva. En el segundo resultó negativa. El médico se pregunta cuál es el valor predictivo positivo en el primer paciente y cuál será el valor predictivo negativo en el segundo.

$$VP+ = \frac{S \times P}{(S \times P) + (1-E) \times (1-P)} = \frac{0.89 \times 0.20}{(0.89 \times 0.20) + (0.10) \times (0.80)} = \frac{0.178}{0.258} = 69\%$$

$$VP- = \frac{E \times (1-P)}{E \times (1-P) + (1-S) \times P} = \frac{0.90 \times (0.80)}{(0.90) \times (0.80) + (0.11) \times (0.20)} = \frac{0.72}{0.742} = 97\%$$

Es decir, si la prueba resulta positiva, la probabilidad estimada de que un individuo realmente esté enfermo es de 0.69 (69%).

*A priori*, la probabilidad de hallarse enfermo es de 0.2 y, al mostrar prueba positiva, esta probabilidad aumenta a 0.69.

Si la prueba resulta negativa en un individuo, la probabilidad estimada de que no esté enfermo es de 0.97 (97%).

El otro método para calcular los valores de predicción se basa en el uso de la prueba de diagnóstico y su comparación con el estándar de oro, en los individuos con sospecha de un padecimiento. El estándar de oro puede ser una prueba de diagnóstico, o bien, la evolución clínica o los resultados de la autopsia. Así, mediante una tabla de  $2 \times 2$ , el grupo de pacientes con pruebas positivas se clasifica en dos grupos: los que tienen el padecimiento (**a** verdaderos positivos) y los que no lo tienen (**b** falsos positivos).

|                          |   | <b>Estándar de oro</b> |   |
|--------------------------|---|------------------------|---|
|                          |   | +                      | - |
| <b>Prueba en estudio</b> | + | a                      | b |
|                          | - | c                      | d |

La proporción de los que tienen el padecimiento es una estimación del valor de predicción de la prueba positiva.

$$VP+ = \frac{a}{a+b}$$

De igual forma, el grupo de pacientes con prueba negativa se clasifica en los dos grupos, los que sí presentaron el padecimiento (**c** falsos negativos) y los que no lo presentaron (**d** verdaderos negativos). La proporción de los que no presentaron el padecimiento es una estimación del valor de predicción de la prueba negativa.

$$VP- = \frac{d}{d+c}$$

Se esquematiza de la siguiente manera:

|                                                          |   | <b>Estándar de oro</b> |                                                                     |
|----------------------------------------------------------|---|------------------------|---------------------------------------------------------------------|
| Individuos<br>positivos con<br>sospecha<br>de enfermedad | P | +                      | a = número de pacientes con la enfermedad.<br>Verdaderos positivos  |
|                                                          | R |                        | b = número de individuos sin la enfermedad.<br>Falsos positivos     |
|                                                          | U |                        | VPP+ = $a/(a+b)$                                                    |
|                                                          | E |                        | d = número de individuos sin la enfermedad.<br>Verdaderos negativos |
|                                                          | B |                        | c = número de individuos con la enfermedad.<br>Falsos negativos     |
|                                                          | A | -                      | VPP- = $d/(d+c)$                                                    |

En el ejemplo anterior, de los 60 pacientes con sospecha de cáncer hepático a los que se les realizó biopsia por aspiración, en 21 se encontró la prueba positiva y en 39 negativa. Más tarde se determinó por medio del estudio histopatológico quiénes realmente tenían cáncer hepático; los resultados fueron los siguientes:

|    |    | <b>Estudio histopatológico</b> |                         |                         |
|----|----|--------------------------------|-------------------------|-------------------------|
| 60 | 21 | biopsia +                      | 17 con CHC<br>4 sin CHC | VPP+ = $17/21 = 81\%$   |
|    | 39 | biopsia -                      | 37 con CHC<br>2 sin CHC | VPP- = $37/39 = 94.8\%$ |

El valor de predicción de la prueba positiva es de 81%. El valor de predicción de la prueba negativa es de 94.8 por ciento.

Estos resultados son válidos para la población de pacientes semejantes a la estudiada, con las mismas técnicas e interpretación de los resultados con los mismos criterios.

## Razones de verosimilitud

El valor predictivo positivo y negativo de una PD es de enorme utilidad para tomar decisiones clínicas, tiene la limitación de que dependen de la prevalencia. Así, los resultados obtenidos no pueden ser utilizados como índices cuando se quiere comparar dos métodos diagnósticos diferentes, ni tampoco extrapolar los resultados a otros estudios o a una población diferente.

Entonces, resulta necesario determinar otros índices de valoración que sean clínicamente útiles y que no dependan de la prevalencia de la enfermedad en una población, sino que sean también aplicables a la población en general.

Mediante la combinación en un solo índice, de la sensibilidad y la especificidad de una prueba diagnóstica, se pueden obtener resultados más confiables sobre su validez. Los índices de verosimilitud, también llamados **razones o cocientes de probabilidad positivos y negativos**, miden cuánto es más probable que un resultado positivo o negativo, según la presencia o ausencia de enfermedad, sea real. Es decir, la razón de verosimilitud indica cómo el resultado de una prueba hará cambiar la probabilidad pretest a la probabilidad posttest de la enfermedad.

Además, como señala Manterola, se pueden obtener varios niveles de una medida, y no necesariamente de forma dicotómica. Otra de sus ventajas es que posibilitan la comparación de diferentes pruebas para un mismo diagnóstico y se mantienen constantes aunque la prevalencia de la enfermedad varíe en las poblaciones en las que se apliquen.

### Razón de verosimilitud positiva

Se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos.

Es el cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivos (1 – especificidad):

$$RV+ = \frac{\text{sensibilidad}}{1 - \text{especificidad}}$$

### Razón de verosimilitud negativa

Se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma.

Es el cociente entre la fracción de falsos negativos (1 – sensibilidad) y la fracción de verdaderos negativos (especificidad):



$$RV = \frac{\text{sensibilidad}}{1 - \text{especificidad}}$$

## Curva ROC

Cuando el resultado de una prueba diagnóstica se expresa en escalas cuantitativas continuas o incluso discretas, es posible estimar indicadores de eficacia a menos que se señale un punto de corte determinado. Así, surgió la necesidad de encontrar un indicador general de eficacia para este tipo de pruebas; la llamada **curva ROC** (*Receiver Operating Characteristic*, o Característica operativa del receptor) brindó esta posibilidad en su aplicación a la Medicina.

Fue después del ataque a Pearl Harbor en 1941 cuando el ejército de Estados Unidos inició un programa de investigación para detectar correctamente los aparatos japoneses a partir de sus señales y ruido de radar. La curva ROC fue desarrollada por ingenieros para medir la eficacia en la detección de objetos enemigos en campos de batalla durante la Segunda Guerra Mundial, mediante pantallas de radar, a partir de lo cual se desarrolló la Teoría de Detección de Señales (TDS).

Esta curva fue propuesta por primera vez para describir la relación entre señal y ruido, y se desarrolló en la comparación de la eficacia de radares. Se necesitaba evaluar la capacidad de un radar para distinguir entre verdaderas señales y ruido de otros tipos. El radar podría equivocarse de dos formas: fallando en la detección de la señal (falso negativo) o detectando una falsa (falso positivo). Así, a los radares se les cambiaba el umbral de detección de señales y este cambio originaba distintas tasas de errores relacionados entre sí: a medida que el umbral disminuía, la tasa de falsos negativos descendía (aumenta la sensibilidad) y aumenta la tasa de falsos positivos (disminuyendo la especificidad).

En el decenio de 1950-1959, la curva ROC se utilizó en psicofísica para evaluar la capacidad de detección de humanos (y también de no humanos) en señales débiles. En Medicina el análisis ROC se ha utilizado de forma muy extensa en Epidemiología e investigación médica. Esta curva se trasladó con facilidad a la evaluación de pruebas diagnósticas, ya que la situación es muy similar a la que le dio origen. Se trata de detectar una enfermedad dada, y la prueba en cuestión puede equivocarse en el sentido de los falsos negativos o los falsos positivos; al cambiar el punto de corte (homólogo del umbral para el radar) cambian las tasas de error, es decir, la sensibilidad y la especificidad.

La **curva ROC** es una representación gráfica de la sensibilidad frente a (1 — especificidad) para un sistema clasificador binario según los diferentes puntos de corte o umbral de discriminación, es decir, valor a partir del cual se acepta que un caso es un positivo. La curva ROC se construye al representar gráficamente en dos ejes de coordenadas (**X**, **Y**), cada uno de los puntos de corte dados por los falsos positivos (1 — especificidad) en el eje de las **X** y la sensibilidad (verdaderos positivos) en el eje de las **Y**.

Un espacio ROC representa los intercambios entre verdaderos positivos y falsos positivos. El mejor resultado posible de predicción o poder de la prueba se sitúa en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de

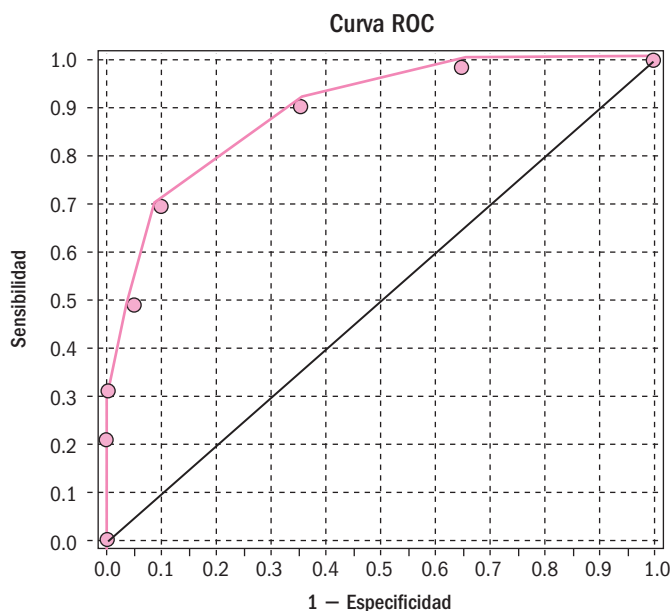


**Figura 12-2.** Curva ROC. Representación gráfica de los dos ejes de coordenadas (X, Y), para distintos puntos de corte. Los falsos positivos ( $1 - \text{especificidad}$ ) corresponden al eje de las X y la sensibilidad (verdaderos positivos) al eje de las Y.

sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo) (figura 12-2).

El área bajo la curva se convierte en el mejor indicador de la capacidad predictiva de la prueba diagnóstica, independientemente de la prevalencia de la enfermedad en la población de referencia.

La curva ROC permite entonces seleccionar el mejor punto de corte y observar comparativamente la sensibilidad y especificidad de una prueba. Además se utilizan para mostrar la relación o comparación entre diferentes valores de sensibilidad y especificidad. Debe tomarse en cuenta que algunos valores cercanos a los límites de normalidad no necesariamente son manifestaciones de enfermedad. Asimismo, es posible que personas que tienen un padecimiento presenten valores normales. Esta superposición exige que se defina un valor crítico que discrimine a los individuos enfermos de los sanos (figura 12-3).

Al definir como valor crítico para el diagnóstico de diabetes tipo 2, en 100, 120 o 130 mg/100 ml de glucosa en sangre en ayunas, se modifican tanto la sensibilidad como la especificidad (figura 12-4).

El comportamiento de dichas pruebas depende de dónde se coloque el punto de corte o “valor crítico”; si éste se desplaza hacia la derecha (valores mayores de glucosa) disminuyen los **falsos positivos**, pero aumentan los **falsos negativos**, o en otros términos, disminuye la sensibilidad y aumenta la especificidad (figuras 12-5 y 12-7).

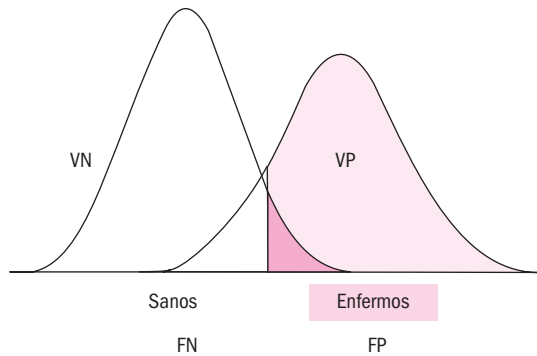


**Figura 12-3.** Superposición entre individuos enfermos de los sanos.

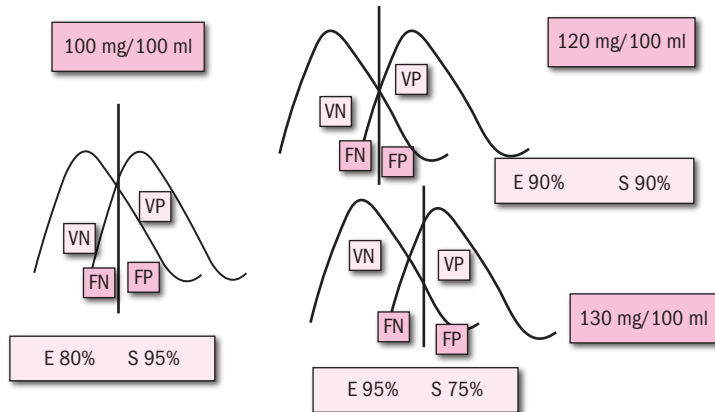


**Figura 12-4.** Sensibilidad y especificidad en valores críticos de 100, 120 o 130 mg/100 ml de glucosa en sangre en ayunas para el diagnóstico de diabetes tipo 2.

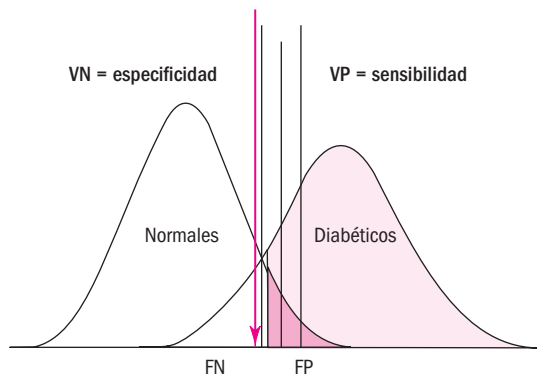


**Figura 12-5.** Punto de corte o "valor crítico" desplazado hacia la derecha (valores mayores de glucosa).

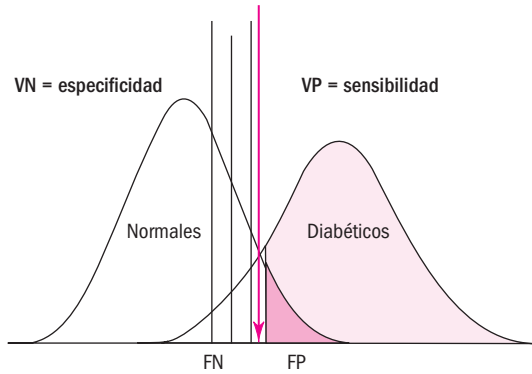


**Figura 12-6.** Punto de corte o “valor crítico” desplazado hacia la izquierda (valores menores de glucosa).

Si se desplaza hacia la izquierda (valores menores de glucosa), disminuyen los **falsos negativos**, pero aumentan los **falsos positivos**, es decir, disminuye la especificidad y aumenta la sensibilidad (figuras 12-6 y 12-7).

### Comparación de pruebas diagnósticas mediante curva ROC

Lo más natural es que la comparación de la eficacia de dos o más pruebas diagnósticas para detectar una enfermedad o proceso patológico dado, pueda hacerse sobre la base de comparar los valores de S y de E de tales pruebas. Pero, cuando se trata de pruebas con resultado cuantitativo, la utilización de la curva ROC correspondiente resulta el modo más adecuado de determinar cuál de las pruebas es más eficaz, puesto que ya se vio que la S y la E de pruebas de este tipo dependen del punto de corte elegido (figura 12-8). Teniendo en cuenta lo que se ha señalado previamente, es posible comprender que la curva ROC que tenga el área mayor será la que corresponda a la prueba más eficaz.

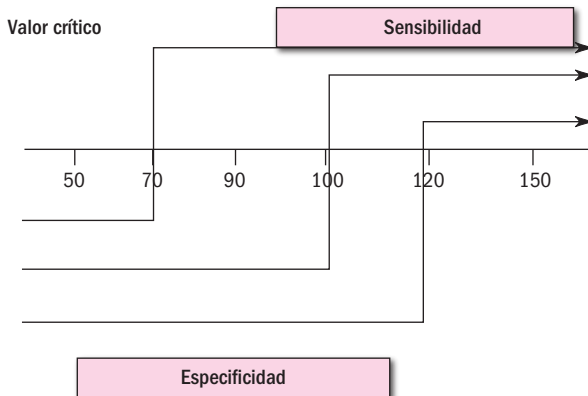


**Figura 12-7.** Diferencias entre la sensibilidad y la especificidad según el punto de corte o valor crítico.

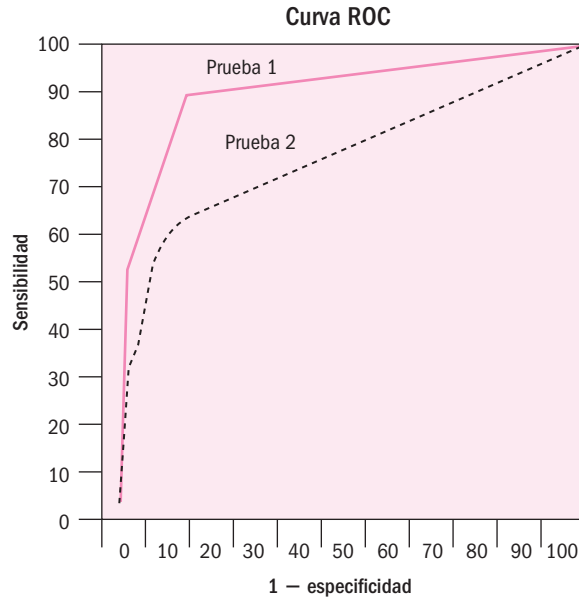


**Figura 12-8.** Curva ROC para la comparación de dos pruebas diagnósticas.

En todo nuevo procedimiento o con el uso de nuevos equipos de diagnóstico, es indispensable realizar cuanto antes la determinación de su valor real en relación con otros métodos, y para elegir la mejor prueba diagnóstica (con mejor sensibilidad y especificidad).

#### Condiciones necesarias para la validación de una prueba diagnóstica

**Características de la población.** La sensibilidad o especificidad de una prueba depende de las características de la población estudiada.

Si se altera o cambia la población en estudio, cambiarán también estos índices. Los datos informados de sensibilidad y especificidad, que son evaluados en poblaciones con una tasa significativa de enfermedad, pueden no ser aplicables en otras poblaciones diferentes en las que se utilice la prueba. Para que este criterio se cumpla, se debe contar con información sobre los siguientes aspectos: sexo y edad de los sujetos en evaluación, resumen de los síntomas clínicos iniciales o estadio de la enfermedad, y criterios de elección para los sujetos que son enrolados en el estudio.

**Subgrupos adecuados.** La sensibilidad y la especificidad pueden representar valores promedio para una población determinada. A menos que el problema para el cual se utiliza la prueba haya sido definido con mucha precisión, aquéllas pueden variar en diferentes subgrupos poblacionales. Para que la prueba pueda ser utilizada con éxito deberían tenerse en cuenta distintos niveles de precisión según los distintos subgrupos

(continúa)

(Continuación)

existentes en la población estudiada. Este criterio se cumple cuando se informa sobre la precisión de la prueba en relación con cualquier subgrupo demográfico o clínico (p. ej., en sujetos sintomáticos y sujetos asintomáticos).

**Sesgo de selección.** Puede producirse cuando los sujetos con los resultados positivos o negativos de una prueba son derivados de forma preferente para verificar el diagnóstico mediante otra prueba considerada el estándar de referencia. Para que este criterio se cumpla, todos los sujetos deberían haber sido asignados para recibir tanto la prueba diagnóstica en estudio como el estándar de referencia a través de un procedimiento directo o mediante el seguimiento clínico.

**Sesgo de medición.** Podría introducirse si la PD o el estándar de referencia se realizan sin tomar precauciones para garantizar la objetividad de su interpretación (similar al enmascaramiento utilizado en los ensayos clínicos para tratamiento). Se puede obviar si la PD en evaluación y el estándar de referencia son interpretados de forma separada y enmascarada por personas independientes que desconocen los resultados de una y otro.

**Precisión de los resultados.** La precisión de la sensibilidad y la especificidad dependen del número de pacientes evaluados. Igual que otras medidas, el resultado estimado debe tener los intervalos de confianza o el error estándar reportados, independientemente de la magnitud encontrada.

**Presentación de resultados indeterminados.** No todas las PD dan lugar a un “sí” o un “no” como respuesta, a veces dan lugar a resultados equívocos o indeterminados. La frecuencia de resultados indeterminados limitará la aplicabilidad de la prueba o la hará más cara si da lugar a otros procedimientos diagnósticos posteriores. La frecuencia de resultados indefinidos y el modo en el que se usan en el cálculo de la precisión de la prueba constituyen una información de importancia crítica para conocer la eficacia de la misma. Para que este criterio se cumpla, el trabajo debe reflejar de forma apropiada todos los resultados positivos, negativos o indeterminados generados durante el estudio, así como si los resultados indeterminados se incluyeron o excluyeron al calcular los indicadores de precisión de la prueba.

**Reproducibilidad de la prueba.** Las pruebas no siempre dan el mismo resultado, por motivos relacionados con la variabilidad de éstas o de la interpretación del observador. Los motivos y el impacto de este asunto deben ser tenidos en cuenta. Para que se cumpla este criterio en pruebas que requieren interpretación del observador, al menos alguna de las pruebas debería ser evaluada con alguna medida que resuma la variabilidad interobservador. Para pruebas sin interpretación del observador, el criterio se cumple cuando se refleja una media que resuma la variabilidad del instrumento.

Manterola C., 2009.

## Pruebas de detección temprana, tamiz o escrutinio

De acuerdo con Jenicek las pruebas de detección representan “la identificación presuntiva de una enfermedad o defecto no reconocido mediante la aplicación de pruebas, exámenes u otros procedimientos, los cuales pueden ser aplicados rápidamente. Las pruebas de detección identifican a las personas en apariencia sanas que probablemente tienen

una enfermedad de aquellas que no la tienen. Una prueba de detección no pretende establecer diagnósticos. Las personas con hallazgos positivos o de sospecha deben ser referidas para que se confirme el diagnóstico y se establezca el tratamiento pertinente”.

Las pruebas de detección se aplican a población asintomática o grupos de población en riesgo para diagnosticar un padecimiento cuya morbilidad y mortalidad pueden disminuirse si la entidad se detecta y trata en forma oportuna. Por ejemplo, el examen más utilizado para identificar CHC es la determinación de alfa fetoproteína (AFP), la que si bien muestra un balance aceptable de sensibilidad y especificidad cuando se determina el punto de corte en 20 ng/ml, su elevación no hace el diagnóstico definitivo. Así, una vez que se ha detectado anomalía en una prueba de detección, en este caso AFP elevada, es necesario realizar pruebas para establecer o descartar el diagnóstico de CHC. Éstas pueden ser, como ya se señaló, biopsia por aspiración guiada por ultrasonido, tomografía computarizada, resonancia magnética, etcétera.

Para decidir si se lleva a cabo una prueba de escrutinio o tamiz han de tomarse en cuenta las siguientes consideraciones:

- El padecimiento debe ser muy frecuente para justificar el enorme esfuerzo que tendrá que hacerse para detectarlo.
- El pronóstico de la entidad tiene que ser lo bastante grave en caso de que se trate en términos de morbilidad, incapacidad que produce, intensidad de las molestias y costo financiero.
- Tiene que haber un periodo suficientemente largo entre el momento en que se puede detectar el padecimiento en individuos asintomáticos y la aparición de los primeros síntomas (tiempo de atención).
- Ha de existir una diferencia importante en el pronóstico del individuo si se lo atiende en la etapa asintomática o si se trata en la fase de aparición de los primeros síntomas.
- La prueba diagnóstica debe tener sensibilidad y especificidad alta; ha de ser fácilmente aplicable, barata y segura, y tiene que ser aceptada tanto por los pacientes como por quienes la aplican.
- Debe existir algún tratamiento eficaz que altere la historia actual del padecimiento.

## Pruebas múltiples

Cuando al realizar estudios a un paciente y resultar todos positivos o negativos, la interpretación es directa, sin embargo, lo más frecuente es que algunas pruebas resulten positivas y otras negativas; entonces la dificultad es mayúscula para establecer un diagnóstico. En el proceso diagnóstico es posible realizar las pruebas en serie o en paralelo; esta decisión depende de la enfermedad en estudio y del paciente a tratar.

**Cuadro 12-2.** Normas generales para la aplicación de los programas de detección.  
Criterios para seleccionar una prueba de detección

| Prueba de detección               | Enfermedad a detectar                                                                                            | Población donde se considera la prueba de detección                        |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Sensible                          | Definida de acuerdo con criterios operativos precisos                                                            | Población altamente expuesta a los factores etiológicos de la enfermedad   |
| Específica                        | Gradiente y espectro de la enfermedad bien especificados                                                         | Prevalencia elevada de la enfermedad a detectar                            |
| Buen valor predictivo             | Alta frecuencia                                                                                                  | Disponibilidad de datos demográficos (denominadores de las tasas)          |
| Reproducible                      | Pronóstico malo                                                                                                  | La detección se perciba como una necesidad y prioridad                     |
| Barata                            | Bien distinguible de la normalidad                                                                               | Disponibilidad de recursos administrativos humanos y materiales            |
| Rápida                            | Claridad en cuanto al curso natural y clínico de la enfermedad                                                   | Disponibilidad de exámenes posdetección para el diagnóstico definitivo     |
| Fácil de aplicar                  | Estado presintomático antes de que el curso clínico sea bien definido                                            | Disponibilidad de recursos para el tratamiento                             |
| Aceptable por los pacientes       | Que exista un tratamiento específico                                                                             | Buenas relaciones entre la población, equipos sanitarios y líderes civiles |
| Definida operativa y técnicamente | Que el tratamiento en una fase de detección precoz dé mejores resultados que si se aplica posteriormente         |                                                                            |
| Validada                          | Que la detección y tratamiento de la enfermedad sea prioridad, entre otras que compiten para programas similares |                                                                            |

Adaptado por L. Moreno A. de Jenicek M, *Epidemiología. La lógica de la epidemiología moderna*, Mason, Barcelona, España, 1996.

A continuación se presentan de manera muy general las ventajas y desventajas de cada una de estas estrategias.

## Pruebas en serie

Se realiza la primera prueba, y si el resultado es positivo, se lleva a cabo la segunda, y así sucesivamente.

- Aumentan la especificidad y el valor predictivo positivo.
- Disminuyen la sensibilidad y el valor predictivo negativo.



### Ventajas

- Se inician con las factibles y pertinentes (menos riesgosas y más baratas).
- Implican menor utilización de laboratorio y gabinete.
- Útiles cuando las disponibles no son muy específicas.
- Mayor seguridad sobre el resultado positivo.

### Limitaciones

- Establecer el diagnóstico conlleva más tiempo.
- Existe probabilidad de no diagnosticar una enfermedad.

### Pruebas en paralelo

Se realizan varias pruebas en forma simultánea y cuando una de estas pruebas es positiva, se considera que existe la enfermedad.

- Aumentan la sensibilidad y el valor predictivo negativo.
- Disminuyen la especificidad y el valor predictivo positivo.

### Ventajas

- El diagnóstico se hace con mayor rapidez.
- Indicadas y útiles en caso de urgencia.
- Existe mayor probabilidad de diagnosticar una enfermedad.
- Útiles cuando las pruebas disponibles no son muy sensibles.

### Limitaciones

- Generan un número importante de falsos positivos (sobrediagnostican).
- Los costos son más elevados.
- Conllevan mayor utilización de laboratorio y gabinete.

A continuación se presentan de manera muy general las ventajas y desventajas de cada una de estas estrategias, mismas que se esquematizan en la figura 12-9.

Finalmente, vale la pena recalcar lo señalado por Laín Entralgo en 1978: “El médico y su criterio, ciencia y conciencia, eran hasta hace poco suficientes para el ejercicio cabal de su ciencia-arte, idea que ha sido rebasada en la actualidad. Las tecnologías tienen su eficacia y eficiencia, hasta cierto punto impersonal, ya no importa quién la realice, sino solamente el estar capacitado para hacerlo bien. Las técnicas se vuelven cada vez más indispensables, pero el médico considerado únicamente como aquel que realiza una técnica, puede ser dispensable y podrá ser desplazado por otro que también la realice. La invasora tecnificación del diagnóstico y del tratamiento, la creciente colectivización de la asistencia médica, el imperativo de la cooperación en equipo, se han concitado para borrar el vínculo entre el paciente y el médico.”

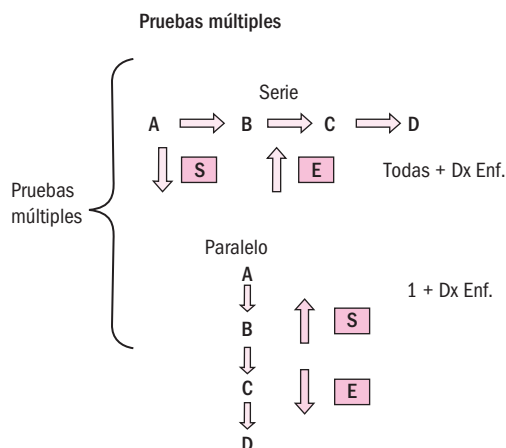


**Figura 12-9.** Sensibilidad y especificidad en las pruebas múltiples.

## Bibliografía

- Bruix J, Sherman M, Llovet JM *et al.* Clinical management of hepatocellular carcinoma: conclusions of the Barcelona-2000 EASL conference. *J Hepatol* 35:421-430. 2001.
- Edmonson HA, Steiner PE. Primary carcinoma of the liver: a study of 100 cases among 48 900 necropsies. *Cancer* 7:462-503. 1954.
- Feinstein A. *Clinical epidemiology: the architecture of clinical research*. Saunders, Filadelfia, EUA, 1985:632-648.
- Ferrel DL *et al.* Proposal for standardized criteria for the diagnosis of benign, borderline, and malignant hepatocellular lesions arising in chronic advanced liver disease. *Am J Surg Pathol* 17(11):1113. 1993.
- Fletcher HR, Fletcher WS, Wagner HE. *Clinical Epidemiology. The essentials*. Williams and Wilkins, Baltimore, EUA. 1982.
- García BC, Moreno Altamirano L, García RH. Evaluación de pruebas diagnósticas. En: Moreno AL, Cano Valle F, García Romero H. *Epidemiología clínica*. McGraw-Hill, UNAM, México. 1994.
- Hirohashi T, Ishak KG, Kojiro M, Wanless IR, *et al.* Pathology and genetics of tumours of the digestive system. En: Hamilton SR, Aalton LA (eds.). *World Health Organization Classification of tumours*. IARC Press, Lyon, Francia, pp. 157-172. 2000.
- Jenicek M. *La lógica de la epidemiología moderna*. Masson, Barcelona, España. 1996.
- Manterola C. Cómo interpretar un artículo sobre pruebas diagnósticas. *Rev Med Clin Condes* 20(5):708-717. 2009.
- Méndez RI, Namihira GD *et al.* *El protocolo de investigación*. Trillas, México. 1987.
- Moreno AL. Validación de pruebas diagnósticas. En: Moreno AL, Cano Valle F. *Epidemiología clínica*. Ed. Facultad de Medicina. UNAM, México. 1988.
- Sacket LD, Hayns RB, Towell P. *Clinical epidemiology: A basic science for clinical medicine*. Little Brown, Toronto, Canadá. 1985.

Yerushalmy J. Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Pub Health Rep* 62:1432-1449. 1947.

Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 6;115(5):654-657. 2007.

## Anexo 1

H. García Barrios

### Casos dudosos

En muchas ocasiones no se puede precisar si un resultado es positivo o negativo. Es frecuente, por ejemplo, que tratándose de pruebas de diagnóstico de imagen, el clínico se enfrente a pruebas dudosas con las que no es posible realizar un diagnóstico y, sin embargo, no pueden considerarse normales. Es posible que dichos estudios dudosos sean de dos clases:

1. Aquellos en que no puede determinarse con precisión si la alteración existe. Por ejemplo, en una serie gastroduodenal quizá se observe depósito de material de contraste que genera una imagen parecida a una úlcera duodenal, pero tal vez no sea posible decidir con seguridad si se trata de una úlcera o si sólo es un artefacto del estudio.
2. Estudios en que es evidente que existe una alteración pero no es posible darle una interpretación precisa. Por ejemplo, en un estudio ultrasonográfico se observa una imagen francamente alterada de un ovario, pero no puede establecerse si se trata de un ovario quístico o de un tumor maligno.

Es claro que los casos dudosos se presentan y no es posible soslayarlos. Ha sido bastante común que estos resultados dudosos se eliminen al analizar los datos de una investigación, pero esto no es correcto. Suponga que a un grupo de 100 personas se aplicó una prueba diagnóstica cuyos resultados fueron los siguientes:

50 individuos con la enfermedad:

Prueba positiva = 5 verdaderos positivos

Prueba negativa = 0 falsos negativos

Prueba dudosa = 45 dudosos

50 sujetos sin la enfermedad:

Prueba negativa = 10 verdaderos negativos

Prueba positiva = 0 falsos positivos

Prueba dudosa = 40 dudosos

Si no se toman en cuenta los casos dudosos, se tiene que  $S = 100\%$  y  $E = 100\%$ , lo cual significa una eficacia muy alta de la prueba, hecho evidentemente falso. Lo ade-

cuando es considerar los resultados dudosos como parte importante de las fórmulas de sensibilidad y especificidad.

Para incluir estos casos, deben clasificarse como dudosos con el padecimiento (Dp). Se da este nombre a las pruebas practicadas a personas con la enfermedad, cuyos resultados no muestran datos suficientes que permitan confirmar la presencia del padecimiento.

En ocasiones, una prueba dudosa puede considerarse como verdadera positiva según la precisión con que el investigador desee hacer un diagnóstico. En caso de que sólo se requiera establecer la presencia de alguna anormalidad, es posible calificar como positiva una imagen que muestre una alteración clara, aunque su naturaleza no pueda determinarse. Por el contrario, cuando mediante su naturaleza no pueda determinarse, o cuando mediante la prueba se desea precisar el tipo de lesión, una imagen como la de la serie gastroduodenal mencionada se considera como verdadera positiva si se acepta que el estudio llevó a diagnóstico una lesión en el estómago. Por otra parte, cuando la prueba no permita determinar si la úlcera es benigna o maligna, y esto interesa particularmente para fijar la efectividad de la prueba, se tiene que calificar como dudosa.

Los casos dudosos sin el padecimiento (Ds) son aquellos cuyas pruebas, aplicadas a personas sin la enfermedad, muestran datos que no permiten descartar la presencia de isquemia del miocardio.

Al incluir estos conceptos en las fórmulas de sensibilidad y especificidad se obtiene lo siguiente:

$$S = \frac{\text{Verdaderos positivos}}{\text{Todos los pacientes con la enfermedad}} \quad \text{es decir,}$$

$$S = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{falsos negativos} + \text{dudosos con el padecimiento}}$$

$$E = \frac{\text{Verdaderos negativos}}{\text{Todos los pacientes sin la enfermedad}} \quad \text{es decir,}$$

$$E = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{falsos positivos} + \text{dudosos sin el padecimiento}}$$

Por ejemplo, se realizó un estudio reciente sobre sensibilidad y especificidad en 107 sujetos con probable colecistitis calculosa, a quienes se realizó intervención quirúrgica (con diagnóstico comprobado); a estos individuos se les practicó previamente colecistografía y ultrasonido, y se encontraron los datos que se anotan a continuación.

En los estudios de colecistografía:

|                           |      |
|---------------------------|------|
| Vp = verdaderos positivos | = 25 |
| Fn = falsos negativos     | = 7  |
| Vn = verdaderos negativos | = 15 |
| Fp = falsos positivos     | = 3  |

$D_p$  = dudosos con el padecimiento = 42

$D_s$  = dudosos sin el padecimiento = 15

Al aplicar las fórmulas:

$$S = \frac{V_p}{V_p + F_n + D_p} = \frac{25}{25 + 7 + 42} = \frac{25}{74} = 0.34$$

$$E = \frac{V_n}{V_n + F_p + D_s} = \frac{15}{15 + 3 + 15} = \frac{15}{33} = 0.45$$

En el estudio de ultrasonido se hallaron los siguientes datos:

$V_p$  = verdaderos positivos = 56

$F_n$  = falsos negativos = 2

$V_n$  = verdaderos negativos = 10

$F_p$  = falsos positivos = 13

$D_p$  = dudosos con el padecimiento = 16

$D_s$  = dudosos sin el padecimiento = 10

Al aplicar las fórmulas:

$$S = \frac{V_p}{V_p + F_n + D_p} = \frac{56}{56 + 2 + 16} = \frac{56}{74} = 0.76$$

$$E = \frac{V_n}{V_n + F_p + D_s} = \frac{10}{10 + 13 + 10} = \frac{10}{33} = 0.30$$

La aplicación de estas fórmulas, que toman en cuenta los casos dudosos, permite conocer de manera más aproximada la verdadera eficacia diagnóstica de la prueba. Si se comparan estas dos pruebas diagnósticas, se puede concluir que el ultrasonido es más sensible que la colecistografía para el diagnóstico de colecistitis calculosa, mientras que la colecistografía es un poco más específica para esta enfermedad.

Los índices de sensibilidad y especificidad de una prueba diagnóstica tienen valores diferentes para cada padecimiento, o sea que es posible que una prueba diagnóstica presente variaciones en su sensibilidad según la enfermedad que se pretende diagnosticar.

## Casos no comprobados

¿Los casos no comprobados deben o no tomarse en cuenta para el cálculo de sensibilidad, especificidad y valores predictivos de una prueba?

Después de practicar un estudio, es frecuente que el diagnóstico no se compruebe y que los individuos se consideren como verdaderos positivos o verdaderos negativos. Los

pacientes con resultados negativos de colecistografía sólo ocasionalmente son operados, y el diagnóstico seguro queda sin comprobarse. Por otra parte, en estudios de ultrasonido y tomografía axial computarizada con imágenes que sugieren metástasis de una neoplasia, los individuos con muy poca frecuencia se someten a laparoscopia, laparotomía o necropsia, y en rigor se desconoce si en estos estudios se trata de un verdadero o falso positivo. Cuando no es posible aclarar esta información, lo recomendable es no incluir en la investigación los casos que no se han comprobado. De otra manera, se falsea la eficacia de la prueba.

Por otra parte, es muy conveniente señalar en un trabajo el criterio de comprobación diagnóstica y si ésta se llevó a cabo antes o después de la aplicación de la prueba diagnóstica en estudio.

La finura de la información que una imagen proporcione puede modificar la sensibilidad y especificidad de una prueba diagnóstica.

Se habla del grado de resolución de los diferentes equipos y de las imágenes que brindan; ello, de acuerdo con la claridad y precisión con que se diferencia la imagen que muestra la lesión y aquella que la rodea.

Un equipo de alta resolución es aquel que permite definir con mayor claridad las estructuras y, consecuentemente, diagnosticar lesiones más pequeñas. Conforme avanza la tecnología y se crean nuevas “generaciones” de equipo, el grado de resolución y la nitidez de las imágenes pueden ir aumentando, lo que incrementará también la eficacia de la prueba y disminuirá los casos dudosos.

Es necesario considerar si estas ventajas (que quizá sólo sean aparentes) están en relación con el costo de los equipos y si es conveniente la adquisición de un aparato nuevo que no modifique de manera sustancial la utilidad del que ya se tiene.

Los grupos de personas o pacientes en que se estudia la eficacia de una prueba (sensibilidad, especificidad y valores de predicción) en relación con un padecimiento, siempre son conjuntos con determinadas características que el investigador debe definir con toda precisión y ha de describir al informar sobre el resultado de sus investigaciones.

Los estudios se pueden realizar en diversos grupos poblacionales con características que el investigador debe definir con toda precisión y ha de describir al informar sobre el resultado de sus investigaciones.

Los estudios se pueden realizar en diversos grupos poblacionales con características diferentes. Por ejemplo, es posible estudiar personas:

- a) Con diagnóstico de padecimiento positivo o negativo comprobado
- b) Con enfermedades similares que requieren diagnóstico diferencial
- c) Con factores de riesgo específicos para la entidad
- d) De la población general
- e) Pertenecientes a un hospital o una sala determinada

Los cuadros 12A-1 y 12A-2 son dos ejemplos de cuadros de  $3 \times 2$  utilizados en algunos de estos casos.

**Cuadro 12A-1.** Sensibilidad y especificidad selectivas de enfermos con padecimiento comprobado y con respecto a otros pacientes que requieren diagnóstico diferencial

|                                                           | Pruebas positivas              | Pruebas negativas              | Pruebas dudosas                       |                         |
|-----------------------------------------------------------|--------------------------------|--------------------------------|---------------------------------------|-------------------------|
| Pacientes con el padecimiento comprobado antes del examen | Número de verdaderos positivos | Número de falsos negativos     | Número de dudosos con el padecimiento | Total de pacientes      |
| Personas sanas                                            | Número de falsos positivos     | Número de verdaderos negativos | Número de dudosos con el padecimiento | Total de personas sanas |
|                                                           | Total de positivos             | Total de negativos             | Total de dudosos                      |                         |

Los resultados en cada caso pueden ser muy diferentes. Si se desea conocer la sensibilidad y especificidad de la prueba en los pacientes que requieren diagnóstico diferencial, ésta se realiza precisamente en tal grupo de sujetos y se indica así el informe de la investigación.

Si la efectividad de la prueba se estudia en la población general para detectar pacientes con o sin determinados factores de riesgo en un estudio de tamiz, el investigador determina con toda precisión las características de la muestra y señala que la sensibilidad y especificidad de la prueba se refieren a dicha población.

Es importante hacer notar que sensibilidad, especificidad y valores de predicción de una prueba son siempre selectivos en lo que se refiere al grupo de estudio. Es necesario que el investigador, en cada ocasión, señale las características del grupo en el que llevó a cabo la investigación, y el porqué de la selección.

**Cuadro 12A-2.** Sensibilidad y especificidad selectivas de enfermos con padecimiento comprobado y con respecto a otros pacientes que requieren diagnóstico diferencial

|                                                                         | Pruebas positivas              | Pruebas negativas              | Pruebas dudosas                       |                                                                                  |
|-------------------------------------------------------------------------|--------------------------------|--------------------------------|---------------------------------------|----------------------------------------------------------------------------------|
| Pacientes con el padecimiento comprobado                                | Número de verdaderos positivos | Número de falsos negativos     | Número de dudosos con el padecimiento | Total de pacientes con el padecimiento comprobado                                |
| Pacientes con otros padecimientos que requieren diagnóstico diferencial | Número de falsos positivos     | Número de verdaderos negativos | Número de dudosos con el padecimiento | Total de pacientes con otros padecimientos que requieren diagnóstico diferencial |
|                                                                         | Total de positivos             | Total de negativos             | Total de dudosos                      |                                                                                  |

# Anexo 2

H. García Romero

## Tamaño de la muestra

Para calcular el número de elementos de una muestra en la que se establece una proporción, se ha utilizado el coeficiente de variación, que expresa la desviación estándar como porcentaje de la proporción. El tamaño de la muestra se busca en las tablas en que se incluye este coeficiente de variación y la proporción estimada.

Es probable que en la práctica, el médico prefiera que los límites de confiabilidad no excedan valores fijos, como pueden ser 2.5, 5 o 10% por encima y por debajo de los valores esperados. El tamaño de la muestra en estos casos se puede encontrar en el cuadro 12A-3, elaborado por García Romero.

**Cuadro 12A-3.** Tamaño de la muestra en estudios de sensibilidad y especificidad de una prueba de diagnóstico

| Proporción esperada (%) | Tamaño de la muestra |          |          |
|-------------------------|----------------------|----------|----------|
|                         | p ± 0.025            | p ± 0.05 | p ± 0.10 |
| 0.95                    | 291                  | 76       | 18       |
| 0.90                    | 552                  | 114      | 35       |
| 0.85                    | 782                  | 204      | 49       |
| 0.80                    | 981                  | 256      | 62       |
| 0.75                    | 1 150                | 300      | 72       |
| 0.70                    | 1 288                | 336      | 81       |
| 0.65                    | 1 422                | 364      | 88       |
| 0.60                    | 1 472                | 384      | 92       |
| 0.55                    | 1 518                | 396      | 95       |
| 0.50                    | 1 562                | 400      | 96       |
| 0.45                    | 1 518                | 396      | 95       |
| 0.40                    | 1 472                | 384      | 92       |
| 0.35                    | 1 422                | 364      | 88       |
| 0.30                    | 1 288                | 336      | 81       |
| 0.25                    | 1 150                | 300      | 72       |
| 0.20                    | 981                  | 256      | 62       |
| 0.15                    | 782                  | 204      | 49       |
| 0.10                    | 552                  | 144      | 35       |
| 0.05                    | 291                  | 76       | 18       |



En este cuadro, los valores de  $p$  pueden sustituirse por los de sensibilidad y especificidad esperados. Si no se tiene ninguna información previa que permita suponer algunos valores aproximados, éstos deben determinarse en un estudio piloto.

Si se desea que al final del estudio los límites de confiabilidad se encuentren entre  $p \pm 0.025$  (2.5% por arriba y abajo del valor encontrado), el tamaño de la muestra ha de buscarse en la columna correspondiente, en la línea horizontal del valor de  $p$ . Lo mismo tiene que hacerse si se hallan 5 o 10% por encima o debajo de los valores encontrados. Si la sensibilidad de una prueba es aproximadamente de  $0.90 \pm 0.05$  (85 y 95%), en el cuadro es posible observar que la muestra debe ser de un mínimo de 144 elementos.

# Capítulo 13

## Estimación del pronóstico de la enfermedad

*Antonio Villa Romero  
Mario Enrique Rendón Macías*



### Introducción

Las interrogantes señaladas en el caso clínico que aparecen en este capítulo pueden tener respuesta cuando se está ante estudios de seguimiento que tienen como finalidad la medición del **pronóstico**. Es quehacer de este tipo de estudios observacionales, el conocimiento de la evolución natural de la enfermedad bajo circunstancias de tratamiento convencional (dado que no hay modificación en su comportamiento por no existir una intervención experimental). En la disciplina de Epidemiología Clínica los estudios que permiten la medición del pronóstico se denominan **diseños de cohorte**.<sup>6</sup>

### Caso clínico

---

Paciente de sexo femenino de 22 años de edad, con antecedente heredofamiliar de relevancia, a saber, madre con artritis reumatoide. Antecedente personal patológico de dos embarazos y dos pérdidas fetales. Es trasladada al servicio de urgencias por un cuadro clínico de 8 h de evolución, caracterizado por estado de semiinconsciencia, afasia global y flacidez muscular en miembros superior e inferior derechos. A la exploración física se aprecia anisocoria, parálisis de nervios trigémino e hipogloso derechos junto con eritema malar, así como hemiplejía derecha. Se hace evidente un estado de deshidratación (++). En ambos miembros inferiores hay presencia de edema (+++). La tensión arterial es 145/90 mmHg, frecuencia cardíaca 110 ×', frecuencia respiratoria 28 ×' y temperatura 37.6°C. Se solicita tomografía axial computarizada (TAC) craneal de urgencia, electrocardiograma (ECG), así como biometría hemática (BH), química sanguínea (QS), pruebas de coagulación, punción lumbar para toma y estudio de líquido cefalorraquídeo (LCR) y examen general de orina. En la TAC se reporta una imagen hiperdensa con pérdida en la diferenciación de la materia blanca-gris y borramiento de surcos en el territorio de la arteria cerebral media izquierda, compatible con un infarto

*(continúa)*

cerebral.<sup>1</sup> En el ECG no se revelan datos de arritmia o isquemia miocárdica. En la BH se reportan plaquetas de 90 000 por mm<sup>3</sup> y el resto de los parámetros citológicos normales. En la QS aparece un valor sérico de creatinina elevada (4.3 mg/100 ml). Los tiempos de protrombina y trombina son normales y el TTP activado está alargado. Los dímeros D están elevados. El LCR es normal y en el examen general de orina se reporta la presencia de hematuria y proteinuria. El manejo inicial de la paciente se instaure mediante la corrección del estado de deshidratación por administración de solución fisiológica y del estado protrombótico con heparina de bajo peso molecular. Se asegura la adecuada permeabilidad de vía aérea por intubación y ventilación mecánica. Se le inicia prednisona a dosis de 60 mg/día.

Ante la sospecha de un diagnóstico probable de lupus eritematoso generalizado se indica que debe realizarse la determinación de anticuerpos antifosfolípidos, y se informan a títulos elevados: anticuerpos anticardiolipinas para isotipos IgG e IgM, así como anticuerpos anti- $\beta_2$ -glucoproteína-I para isotipos IgG e IgM, así como prueba VDRL positiva y células LE positivas. Otros anticuerpos reportados como positivos fueron anticuerpos antinucleares y anticuerpos anti-DNA de doble cadena.

Debido a los antecedentes y cuadro clínico, así como por los reportes de laboratorio e imagen, se establece el diagnóstico de lupus eritematoso generalizado (al tener  $\geq 4$  criterios ACR-1982),<sup>1-2</sup> con síndrome antifosfolípido secundario, enfermedad renal crónica e infarto cerebral.<sup>3-5</sup>

La paciente se estabiliza clínicamente y evoluciona en forma satisfactoria al cuadro inicial. Continúa su manejo mediante la administración de esteroide y anticoagulante. Se le indica biopsia renal para determinar el grado y extensión de la glomerulonefritis.

Las siguientes preguntas podrían ser derivadas del caso clínico anterior:

- ¿Cuál es el pronóstico para la supervivencia a 5 y 10 años de una paciente de 22 años de edad con diagnóstico de lupus eritematoso generalizado?
- ¿Cuál es la probabilidad de que esta paciente presente un segundo evento trombotico cerebral dado que ya tuvo uno previo?
- ¿Cuál es el riesgo de llegar a enfermedad renal crónica terminal?
- ¿Cuál es la probabilidad de que la paciente pueda tener un producto vivo de término si ya tuvo pérdidas fetales repetidas como parte del síndrome antifosfolípido secundario?

El término “cohorte” en Epidemiología Clínica se refiere a un grupo de personas que comparten una característica en común y que son seguidas a través del tiempo.

De esta forma, en estudios de pronóstico la definición de **cohortes de seguimiento** es fundamental. A través de la experiencia observada en el comportamiento de eventos incidentes y muertes asociadas, se podrán realizar estimaciones sobre la probabilidad de ocurrencia. Asimismo, se podrán realizar comparaciones entre cohortes para establecer la fuerza de asociación de los factores pronósticos.<sup>7</sup>

Por otra parte, en los estudios de seguimiento es propósito fundamental establecer la diferencia entre **factor de riesgo** y **factor pronóstico**.<sup>8</sup>

El **factor de riesgo** es aquel elemento, generalmente expresado como una exposición ambiental o determinante genético, que se asocia con una probabilidad de inicio del proceso etiológico (figura 13-1).

El **factor pronóstico** es aquel elemento, generalmente expresado como característica subclínica del paciente, que se asocia con una mayor o menor probabilidad de inicio del cuadro clínico o de alteración en el desenlace final de la enfermedad (figura 13-1).

En el lupus eritematoso generalizado, por ejemplo, el antecedente familiar de enfermedades reumáticas autoinmunes como artritis reumatoide, síndrome de Sjögren primario, esclerodermia, dermatopolimiositis, además del lupus mismo, constituyen **factores de riesgo** para el desarrollo de la enfermedad en una mujer clínicamente sana.<sup>10</sup> Por otra parte, la elevación de títulos de anticuerpos anticardiolipina o anti- $\beta_2$ -glucoproteína-I constituye un **factor de pronóstico** para el desarrollo del síndrome antifosfolípido secundario en lupus, además que se asocia con una alta probabilidad de presentar eventos trombóticos arteriales o venosos como la pérdida fetal (por trombosis placentaria) o infarto cerebral (por trombosis arterial de vasos cerebrales mayores). La nefropatía por lupus es un **factor de pronóstico** para la sobrevida relacionada con la evolución a enfermedad renal crónica terminal y muerte.<sup>11-20</sup>

De esta forma, tanto el **factor de riesgo** como el **factor de pronóstico** están asociados con una probabilidad de ocurrencia de un evento. Para estimar dicha probabilidad se requiere de la definición de un periodo en el tiempo:  $t_0 - t_k$ . Es decir, es necesario delimitar en el tiempo, un inicio ( $t_0$ ) y un fin ( $t_k$ ). Además, se requiere tener la certeza de que al inicio del periodo ( $t_0$ ), la cohorte en estudio se encuentra libre del evento incidente a medir (ausencia del evento en forma prevalente en  $t_0$ ).

Otro concepto fundamental en estudios de pronóstico lo constituye el **tiempo en riesgo**, mismo que representa la cuantificación a través de la suma de tiempos (en días, meses, años, etc.) de aquellas personas que permanecen bajo observación en el seguimiento y contribuyen con tiempos específicos desde que ingresan a la cohorte y hasta que presentan el evento incidente o muerte en estudio, se pierden durante el seguimiento o llegan al final de este último.<sup>9</sup>

## Medición de eventos incidentes de pronóstico<sup>9,21</sup>

Suponga una cohorte hipotética de 10 pacientes de sexo femenino con diagnóstico de lupus eritematoso generalizado, las cuales son seguidas en el tiempo por un máximo de cinco años para medir la incidencia de síndrome antifosfolípido secundario y el pronóstico de mortalidad asociada con dicho síndrome (figura 13-2).

De tal forma, la representación de esta cohorte en el tiempo permite señalar lo siguiente:

- La paciente N° 1 tiene diagnóstico de lupus (como todas las pacientes de la cohorte) y fue seguida durante un año y medio antes de que se le estableciera el diagnóstico de síndrome antifosfolípido. Estuvo bajo vigilancia médica durante dos años más y se perdió el seguimiento en el año 3.5.

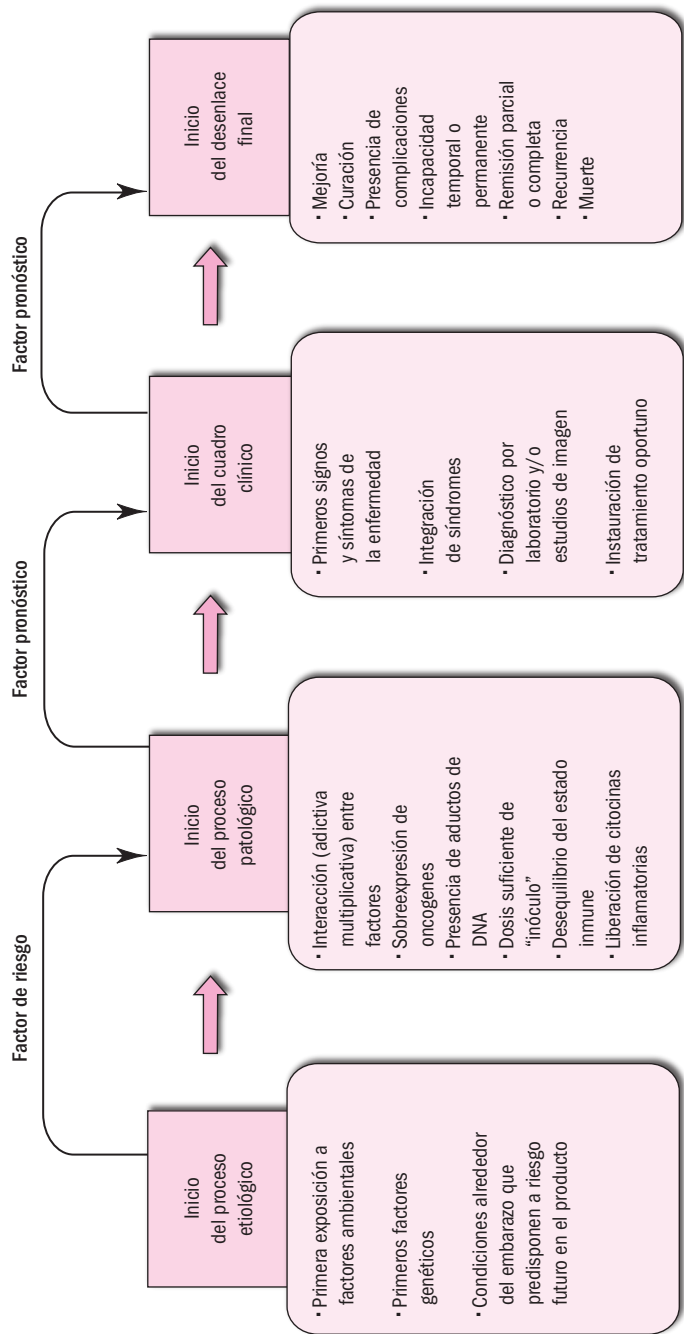


**Figura 13-1.** Ubicación de los factores de riesgo y de pronóstico dentro de la evolución natural de la enfermedad. (Modificada de: Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Lifetime Learning Publications, Belmont, EUA, 1982.)<sup>9</sup>

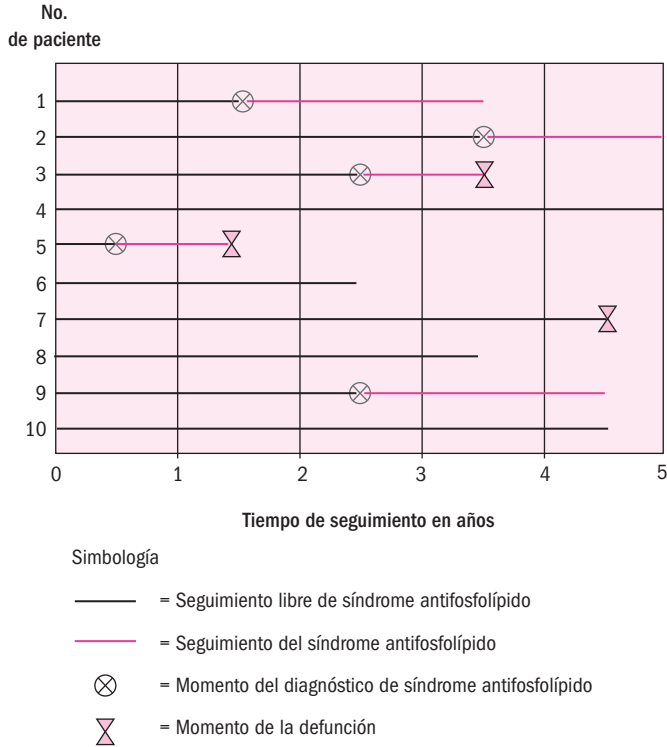


**Figura 13-2.** Representación de una cohorte hipotética de 10 pacientes con lupus eritematoso generalizado seguida por un máximo de cinco años para medir incidencia de síndrome antifosfolípido secundario y mortalidad.

- Así, la paciente N° 2 es seguida en vigilancia clínica durante 3.5 años. Se le diagnostica síndrome antifosfolípido en ese tiempo y recibe vigilancia 1.5 años más hasta el cierre del seguimiento del estudio (año 5).
- La paciente N° 3 es seguida 2.5 años, se le diagnostica síndrome antifosfolípido y es monitorizada por un año más hasta que fallece a causa de complicaciones por este síndrome.
- La paciente N° 4 recibió seguimiento durante cinco años (todo el periodo del estudio) y nunca desarrolló síndrome antifosfolípido.
- La paciente N° 5 estuvo bajo supervisión medio año, desarrolló síndrome antifosfolípido, fue seguida un año más y falleció a consecuencia de complicaciones del síndrome.
- La paciente N° 6 fue seguida durante 2.5 años y no desarrolló síndrome antifosfolípido durante ese lapso. Más adelante dejó de estar bajo supervisión clínica.
- La paciente N° 7 estuvo bajo vigilancia médica durante 4.5 años hasta el momento de su muerte. Durante ese plazo no desarrolló síndrome antifosfolípido, por lo que la causa de su muerte no se puede atribuir a alguna complicación del síndrome.

- La paciente N° 8 fue vigiada durante 3.5 años, para después dejar de estar bajo supervisión. No desarrolló síndrome antifosfolípido.
- La paciente N° 9 fue seguida por 2.5 años y se le diagnosticó síndrome antifosfolípido. Fue monitorizada por dos años más y se perdió seis meses antes de completar el seguimiento total del estudio.
- La paciente N° 10 fue seguida durante 4.5 años y se perdió antes de completar el seguimiento total del estudio. No desarrolló síndrome antifosfolípido.

Con base en la información previa, es factible calcular medidas epidemiológicas que permitan responder a las siguientes preguntas:

**Pregunta 1.** ¿Qué probabilidad tiene una paciente de esta cohorte de lupus eritematoso generalizado de desarrollar síndrome antifosfolípido secundario en el periodo  $t_0 - t_5$ ?

**Respuesta 1.** Estimar la incidencia acumulada (IncAc) (en inglés, *cumulative incidence*) de síndrome antifosfolípido en el periodo:  $t_0 - t_5$ :

$$\text{IncAc } t_0 - t_5 = \frac{\text{Número casos incidentes del síndrome antifosfolípido}}{\text{Pacientes sin diagnóstico del síndrome antifosfolípido en } t_0}$$

En el ejemplo:

$$\text{IncAc } t_0 - t_5 = \frac{5}{10} = 0.5$$

**Aclaración 1.** El denominador de esta medida se basa en la población libre del evento incidente (síndrome antifosfolípido) al inicio del periodo de seguimiento. El resultado constituye una probabilidad (es decir, un valor entre 0 y 1) que generalmente se multiplica por 100 para expresarlo en porcentaje. **Probabilidad y riesgo son términos equivalentes en Epidemiología Clínica.**

**Interpretación 1.** La probabilidad o riesgo que tiene una paciente de esta cohorte de lupus de desarrollar (ser un caso nuevo o incidente) síndrome antifosfolípido en el periodo  $t_0 - t_5$  es igual al 50%.

**Pregunta 2.** ¿Qué probabilidad tiene una paciente de esta cohorte de lupus de fallecer por complicaciones del síndrome antifosfolípido secundario en el periodo  $t_0 - t_5$ ?

**Respuesta 2.** Estimar la mortalidad acumulada (MorAc) (en inglés: *cumulative mortality*) por complicaciones del síndrome antifosfolípido en el periodo  $t_0 - t_5$ :

$$\text{MorAc } t_0 - t_5 = \frac{\text{Número de muertes en presencia del síndrome antifosfolípido}}{\text{Pacientes vivas con o sin diagnóstico del síndrome antifosfolípido en } t_0}$$

En el ejemplo:

$$\text{MorAc } t_0 - t_5 = \frac{2}{10} = 0.2$$

**Aclaración 2.** La paciente N° 7 no se contabiliza debido a que falleció por otra causa distinta a complicaciones del síndrome antifosfolípido. El denominador es 10 (todas las pacientes de la cohorte) debido a que todas ellas estuvieron libres de síndrome antifosfolípido al inicio del periodo ( $t_0$ ) y durante el seguimiento de cinco años pudieron desarrollar el síndrome y, por tanto, morir por complicaciones de éste.

**Interpretación 2.** La probabilidad o riesgo que tiene una paciente de esta cohorte de lupus de fallecer por complicaciones del síndrome antifosfolípido en el periodo  $t_0 - t_5$  es igual al 20%.

**Pregunta 3.** ¿Cuál es la tasa o velocidad de incidencia del síndrome antifosfolípido en el periodo  $t_0 - t_5$ ?

**Respuesta 3.** Estimar la densidad de incidencia (DenInc) (del inglés, *hazard function*, *hazard rate* o *incidence density*) del síndrome antifosfolípido en el periodo  $t_0 - t_5$ .

$$\text{DenInc } t_0 - t_5 = \frac{\text{Número de casos incidentes del síndrome antifosfolípido}}{\text{Suma de años de seguimiento libre del síndrome antifosfolípido}}$$

Suma de años de seguimiento libre de síndrome antifosfolípido

En el ejemplo:

$$\text{DenInc } t_0 - t_5 = \frac{5}{30.5 \text{ años-persona}} = 0.16 \text{ casos incidentes por año-persona}$$

**Aclaración 3.** Para el denominador de esta medida, debe realizarse la suma de los años de seguimiento con los cuales contribuye cada paciente hasta el momento en que se le diagnostica síndrome antifosfolípido o se pierde al seguimiento o fallece por una causa distinta. Este denominador se conoce como tiempo-poblacional, que generalmente se expresa como años-persona de seguimiento. Para facilitar la interpretación, el resultado final (que es por cada año-persona) se multiplica por cualquier múltiplo de 10 hasta llevarlo a números enteros.

**Interpretación 3.** La velocidad o densidad con la que se observan casos incidentes de síndrome antifosfolípido en esta cohorte durante el periodo  $t_0 - t_5$  es igual a 16 casos nuevos por cada 100 años-paciente de seguimiento.

**Pregunta 4.** ¿Cuál es la tasa o velocidad de mortalidad por complicaciones del síndrome antifosfolípido en el periodo  $t_0 - t_5$ ?

**Respuesta 4.** Estimar la densidad de mortalidad (DenMor) (del inglés, *hazard function*, *hazard rate* o *mortality density*) de complicaciones del síndrome antifosfolípido en el periodo  $t_0 - t_5$ .

$$\text{DenMor } t_0 - t_5 = \frac{\text{Número de muertes por complicaciones del síndrome antifosfolípido}}{\text{Suma de años de seguimiento con diagnóstico de síndrome antifosfolípido hasta la muerte o pérdida o fin de seguimiento}}$$



En el ejemplo:

$$\text{DenMor } t_0 - t_5 = \frac{2}{7.5 \text{ años-persona}} = 0.27 \text{ muertes por año-persona}$$

**Aclaración 4.** Para el denominador de esta medida, deberá realizarse la suma de los años de seguimiento con los cuales contribuye cada paciente al riesgo de muerte por complicaciones del síndrome antifosfolípido o hasta la pérdida al seguimiento o muerte por una causa distinta. Para estimar esta medida, sólo se tomarán los tiempos en riesgo desde el momento en que se establece el diagnóstico de Síndrome Antifosfolípido (líneas verdes en la figura 13-2). Es decir, aunque resulte obvio, para morir por complicaciones del síndrome antifosfolípido hay que tener síndrome antifosfolípido.

**Interpretación 4.** La velocidad o densidad con la que se observan muertes por complicaciones del síndrome antifosfolípido en esta cohorte durante el periodo  $t_0 - t_5$  es igual a 27 muertes por cada 100 años-paciente de seguimiento.

## Medición de la fuerza de asociación en estudios de pronóstico<sup>9,21</sup>

Suponga que en un estudio clínico de pacientes con lupus eritematoso generalizado se decide establecer una comparación entre dos cohortes de pacientes: aquellas con nefropatía (cohorte con factor de pronóstico: **conFP**) y aquellas sin nefropatía (cohorte sin factor de pronóstico: **sinFP**). En ambas se establece un seguimiento de 10 años en el tiempo para medir y comparar el riesgo de muerte (**M**) o sobrevida (**S**), como se muestra en el recuadro siguiente.

|              | <b>M</b>     | <b>S</b>     |                      |
|--------------|--------------|--------------|----------------------|
| <b>conFP</b> | <i>a</i>     | <i>b</i>     | <i>a + b</i>         |
| <b>sinFP</b> | <i>c</i>     | <i>d</i>     | <i>c + d</i>         |
|              | <i>a + c</i> | <i>b + d</i> | <i>a + b + c + d</i> |

Si las celdas **a** y **c** corresponden a las muertes observadas durante el seguimiento en las cohortes **conFP** y **sinFP**, respectivamente, entonces la mortalidad acumulada se calcula:

$$\text{Mortalidad acumulada en la cohorte } \text{conFP} = a/(a + b)$$

Esta medida determinará el riesgo de morir de una paciente con nefropatía en el periodo de 10 años.

$$\text{Mortalidad acumulada en la cohorte } \text{sinFP} = c/(c + d)$$

Esta medida determinará el riesgo de morir de una paciente sin nefropatía en el periodo de 10 años.

Al establecer una comparación entre las dos mortalidades acumuladas, se obtiene la medida de asociación conocida como **riesgo relativo (RR)**:

$$\frac{a/(a+b)}{c/(c+d)} = \text{RR (riesgo relativo)}$$

Al corresponder a una razón aritmética, existen tres posibilidades de resultado del RR:

RR > 1 que representa un mayor riesgo de daño (enfermedad o muerte) en la cohorte **conFP** que en la cohorte **sinFP**

RR = 1 que representa el mismo riesgo de daño (enfermedad o muerte) en la cohorte **conFP** que en la cohorte **sinFP**

RR < 1 que representa un menor riesgo de daño (enfermedad o muerte) en la cohorte **conFP** que en la cohorte **sinFP**

Ejemplo:

**Pregunta 1.** ¿Qué probabilidad tiene una paciente con lupus eritematoso generalizado y nefropatía de fallecer durante el periodo de 10 años de seguimiento?

**Respuesta 1.** Estimar la mortalidad acumulada en la cohorte con nefropatía (**conFP**) =  $a/(a+b) = 100/500 = 0.2$ .

**Interpretación 1.** La probabilidad o riesgo que tiene una paciente con lupus eritematoso generalizado y nefropatía de morir en 10 años de seguimiento es igual a 20%.

**Pregunta 2.** ¿Qué probabilidad tiene una paciente con lupus eritematoso generalizado y sin nefropatía de fallecer durante el periodo de 10 años de seguimiento?

**Respuesta 2.** Estimar la mortalidad acumulada en la cohorte sin nefropatía (**sinFP**) =  $c/(c+d) = 50/500 = 0.1$ .

**Interpretación 2.** La probabilidad o riesgo que tiene una paciente con lupus eritematoso generalizado sin nefropatía de morir en 10 años de seguimiento es igual a 10%.

**Pregunta 3.** ¿Cuántas veces es más probable que muera en un periodo de 10 años una paciente con lupus eritematoso generalizado y nefropatía en comparación con una paciente con lupus eritematoso generalizado sin nefropatía?

**Respuesta 3.** Calcular el riesgo relativo (RR):

$$\text{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{100/500}{50/500} = 2$$

**Interpretación 3.** Una paciente con lupus eritematoso generalizado y nefropatía tiene dos veces el riesgo de morir en comparación con una paciente con lupus eritematoso generalizado sin nefropatía.

También se puede expresar como un riesgo 100% mayor en la cohorte de pacientes **conFP** (RR = 2 menos 1 y el resultado multiplicado por 100).

Ahora suponga que puede estimar y sumar de todas las pacientes de las cohortes (**conFP** y **sinFP**) los tiempos en riesgo (**suma de tiempos en riesgo = STR**) hasta el momento de la muerte, hasta el momento de la pérdida al seguimiento o hasta el fin del estudio.

|              | <i>M</i>     | <i>S</i>     |                      |
|--------------|--------------|--------------|----------------------|
| <i>conFP</i> | <i>a</i>     | <i>b</i>     | <i>a + b</i>         |
| <i>sinFP</i> | <i>c</i>     | <i>d</i>     | <i>c + d</i>         |
|              | <i>a + c</i> | <i>b + d</i> | <i>a + b + c + d</i> |

Donde:

*M* = muertes

*STR* = suma de tiempos en riesgo

*conFP* = cohorte de pacientes con el factor pronóstico

*sinFP* = cohorte de pacientes sin el factor pronóstico

*a* = número de muertes en el periodo en la cohorte *conFP*

*c* = número de muertes en el periodo en la cohorte *sinFP*

*TRcon* = suma de tiempos en riesgo en la cohorte *conFP*

*TRsin* = suma de tiempos en riesgo en la cohorte *sinFP*

*a + c* = total de muertes en el periodo

*TRtot* = suma total de tiempos en riesgo

De esta forma, será posible calcular la función o densidad de mortalidad (*hazard function* o *hazard rate*) en la cohorte *conFP*, en la cohorte *sinFP* o en forma total.

*DenMor conFP* =  $a/TRcon$

*DenMor sinFP* =  $c/TRsin$

*DenMor total* =  $(a + c)/TRtot$

Asimismo, será posible establecer una comparación entre la densidad de mortalidad en la cohorte *conFP* y la cohorte *sinFP*. Esta medida de asociación se denomina **razón de densidades de mortalidad** (RDM) o más conocida por el término en inglés **hazard ratio** (HR).

$$RDM = HR = \frac{a/TRcon}{c/TRsin}$$

Ejemplo:

|              | <i>M</i> | <i>STR</i><br><i>años-paciente</i> |
|--------------|----------|------------------------------------|
| <i>conFP</i> | 100      | 3 500                              |
| <i>sinFP</i> | 50       | 5 000                              |
|              | 150      | 8 500                              |

**Pregunta 1.** ¿Cuál es la velocidad con la que se presentan muertes en las pacientes con lupus eritematoso generalizado y nefropatía?

**Respuesta 1.** Estimar la densidad de mortalidad en la cohorte *conFP*:

*DenMor conFP* =  $a/TRcon = 100/3\,500 = 0.02857$  por año-paciente

**Interpretación 1.** La densidad o velocidad de muertes en la cohorte de pacientes con lupus eritematoso generalizado y nefropatía es = 28 muertes por cada 1 000 años-paciente de seguimiento.

**Pregunta 2.** ¿Cuál es la velocidad con la que se presentan muertes en las pacientes con lupus eritematoso generalizado y sin nefropatía?

**Respuesta 2.** Estimar la densidad de mortalidad en la cohorte *sinFP*:

$$DenMor\ sinFP = c/TRsin = 50/5\ 000 = 0.01\ por\ año-paciente$$

**Interpretación 2.** La densidad o velocidad de muertes en la cohorte de pacientes con lupus eritematoso generalizado y sin nefropatía es = 10 muertes por cada 1 000 años-paciente de seguimiento.

**Pregunta 3.** ¿Cuántas veces es mayor la densidad de mortalidad en la cohorte de pacientes *conFP* en comparación a la cohorte de pacientes *sinFP*?

**Respuesta 3.** Estimar la razón de densidades de mortalidad o *hazard ratio*:

$$RDM = HR = \frac{a/TRcon}{c/TRsin} = \frac{100/3\ 500}{50/5\ 000} = 2.857$$

**Interpretación 3.** La densidad o velocidad de muertes en la cohorte de pacientes *conFP* es 2.8 veces la correspondiente a la cohorte de pacientes *sinFP*.

## Aplicación del análisis de sobrevida en estudios de pronóstico<sup>22</sup>

Sin lugar a dudas, una de las aplicaciones fundamentales del análisis de sobrevida es en los estudios de pronóstico. El análisis de sobrevida está basado en métodos actuariales de cálculo de probabilidades acumuladas de sobrevida a través del tiempo.

El objetivo central en el análisis de sobrevida es estimar la *función acumulada de sobrevida* a través del tiempo. La denominación es la siguiente:

$$\hat{S}_t = \text{estimador de la función acumulada de sobrevida a un tiempo } t$$

Al iniciar un seguimiento, la probabilidad de estar vivo es del 100%. Conforme el tiempo transcurre, dicha probabilidad irá disminuyendo hasta que tienda a ser aproximadamente de cero.

Por tanto, los límites teóricos son:  $\hat{S}_0 = 1$  y  $S_\infty \approx 0$  (figura 13-3).

Entre los conceptos que deben ser comprendidos de manera adecuada para calcular el  $\hat{S}_t$  se encuentra el fenómeno de censuramiento (en inglés: *censuring*).<sup>22</sup>

- El **censuramiento** consiste en la posibilidad de no observar los **tiempos de falla** (terminación) en todos los sujetos.
- Una **falla** representa la ocurrencia del evento en estudio: muerte, caso incidente, remisión, recaída, etcétera.
- El censuramiento ocurre cuando está disponible información incompleta acerca del tiempo de sobrevida de algunos sujetos.

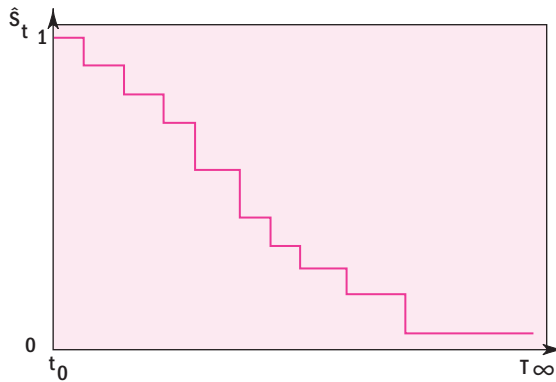


**Figura 13-3.** Disminución del estimador de la función acumulada de supervivencia a través del tiempo.

De esta forma, se consideran:

- **Observaciones no censuradas:** cuando se completa el seguimiento por falla o término del estudio.
- **Observaciones censuradas:** cuando hay pérdida al seguimiento o muerte por una causa diferente a la de estudio.

En el análisis de supervivencia se deben considerar los siguientes supuestos centrales:<sup>22</sup>

1. Los mecanismos de **falla** (término del estudio) y **censuramiento** son estadísticamente independientes.
2. El sujeto que es **censurado** en el tiempo  $t_i$  puede representar a los sujetos que sobreviven  $t_i$ .
3. El **pronóstico** de un sujeto que ha sobrevivido a  $t_i$  no se modifica si el sujeto es **censurado** en  $t_i$ .

### Método de Kaplan-Meier para el cálculo de $\hat{S}_t$ <sup>22,23</sup>

Uno de los métodos más conocidos y aplicados para el cálculo de  $\hat{S}_t$  es el de Kaplan-Meier, originalmente descrito en 1958:<sup>23</sup>

Dicho método consiste en:

1. Ordenar los tiempos  $t_i$  de menor a mayor, “arrastrando” la información de censuramiento.
2. Definir ( $z$ ) de modo que cada intervalo contenga sólo una observación (si no hay empate entre observaciones).
3. Si hay empate y ocurre entre observaciones censuradas, no se modifica el cálculo.
4. Si el empate ocurre entre una observación censurada y una no censurada, se debe asumir que las observaciones censuradas estarán inmediatamente después (a la derecha) de las no censuradas, en el ordenamiento de los datos.

5. Si el empate ocurre entre observaciones no censuradas, se repartirá infinitesimalmente, de modo que  $1 - d_i/n_i$ .

Sea:  $d_i$  = número de fallas entre  $t_{i-1}$  y  $t_i$

$e_i$  = número de censuras entre  $t_{i-1}$  y  $t_i$

$n_i$  = número de sujetos en riesgo al inicio del intervalo entre  $t_{i-1}$  y  $t_i$

**Fórmula para estimar la función acumulada de sobrevivida por el método de Kaplan-Meier.**

$$\hat{S}_{t-KM} = \prod (1 - [d_i/n_i])$$

**Fórmula para calcular intervalos de confianza al 95% del estimador de la función acumulada de sobrevivida por el método de Kaplan-Meier.**

$$(\hat{S}_{t-KM})(e^{\pm 1.96 \sqrt{\Sigma(d_i/n_i(n_i - d_i))}})$$

Ejemplo:

Regresando a la información de la figura 13-1, la cual incluye datos de una cohorte hipotética de 10 pacientes con lupus eritematoso generalizado, podemos señalar los siguientes tiempos en riesgo para mortalidad (**por cualquier causa, con o sin síndrome antifosfolípido**) de la paciente N° 1 a la paciente N° 10:

3.5+, 5+, 3.5, 5+, 1.5, 2.5+, 4.5, 3.5+, 4.5+, 4.5+

Observe que a la derecha de una observación censurada se coloca el símbolo “+” para diferenciarla de las observaciones no censuradas.

Paso 1. Ordenar los tiempos NO censurados de mayor a menor:

1.5, 2.5+, 3.5, 3.5+(2), 4.5, 4.5+(2), 5+(2)

Observe que a la derecha del tiempo censurado o no censurado se señala entre paréntesis el número de empates.

Paso 2. Calcular para cada tiempo no censurado el  $\hat{S}_i$

| Tiempo en riesgo<br>no censurado<br>y censurado             | 0                   | 1.5                    | 2.5+                | 3.5               | 3.5+                | 4.5                   | 4.5+                | 5.0+                |
|-------------------------------------------------------------|---------------------|------------------------|---------------------|-------------------|---------------------|-----------------------|---------------------|---------------------|
| $n_i$ = (sujetos en<br>riesgo al inicio<br>del tiempo $i$ ) | 10                  | 10                     | 9                   | 8                 | 7                   | 5                     | 4                   | 2                   |
| $d_i$ = (fallas)                                            | 0                   | 1                      | 0                   | 1                 | 0                   | 1                     | 0                   | 0                   |
| $e_i$ = (censuras)                                          | 0                   | 0                      | 1                   | 0                 | 2                   | 0                     | 2                   | 2                   |
| $1 - (d_i/n_i)$                                             | 1 –<br>(0/0)<br>= 1 | 1 –<br>(1/10)<br>= 0.9 | 1 –<br>(0/9)<br>= 1 | 1 –<br>(1/8)<br>= | 1 –<br>(0/7)<br>= 1 | 1 –<br>(1/5)<br>= 0.8 | 1 –<br>(0/5)<br>= 1 | 1 –<br>(0/2)<br>= 1 |
|                                                             |                     |                        |                     | 0.875             |                     |                       |                     |                     |
| $\hat{S}_i = \prod (1 - [d_i/n_i])$                         | 1                   | 0.9                    | 0.9                 | 0.7875            | 0.7875              | 0.63                  | 0.63                | 0.63                |

Observe que en tiempos censurados (+) el estimador de la función de supervivencia ( $\hat{S}_i$ ) no cambia, por lo que se puede hacer el cálculo con base únicamente en tiempos no censurados.

Paso 3. Graficar los valores del último renglón de la tabla previa (figura 13-4).

Observe que las “caídas” en los peldaños de la “escalera” están definidas por los cambios en el estimador de la función acumulada de supervivencia en los **tiempos no censurados**.

Paso 4. Interpretar la curva de supervivencia:

La probabilidad que tiene una paciente de esta cohorte de sobrevivir a cinco años del seguimiento es del 63%.

## Sesgos en estudios de pronóstico

Como en todos los estudios clínicos, los estudios de pronóstico pueden generar conclusiones erróneas asociadas a deficiencias en la selección, seguimiento o evaluación de los participantes. En los estudios de pronóstico los principales sesgos de selección suelen ser causados por la inclusión de pacientes con diferentes tiempos de evolución (estadios o progresión de la enfermedad), dado que el pronóstico de una enfermedad depende mucho del momento en el cual ésta se detecte. Para poder comparar la evolución entre diferentes enfermos, por tanto, es necesario agruparlos con base en un mismo momento clínico de ésta; esto puede ser: al diagnóstico, al inicio o terminación de un tratamiento, al momento de la presentación de una complicación o secuela, etcétera. Este momento se considera en la cohorte de estudio como “tiempo cero” o “cohorte de inicio”.

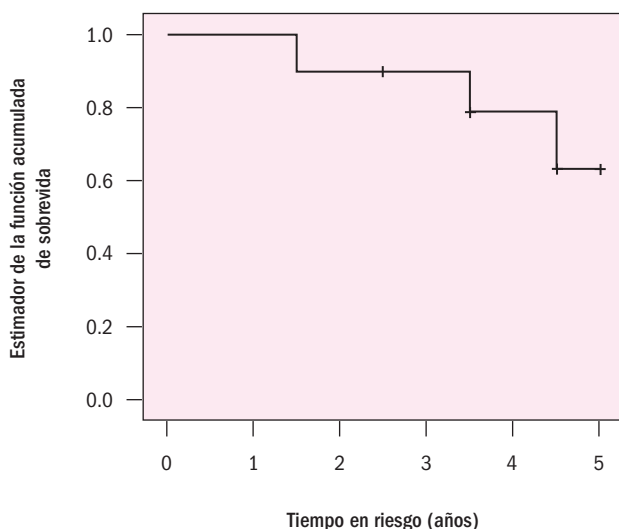


**Figura 13-4.** Estimador de la función acumulada de supervivencia por el método de Kaplan-Meier para los datos de la cohorte hipotética de la figura 13-1.

Cuando los sujetos analizados en un estudio tienen diferentes tiempos “cero” o “de inicio”, existe la posibilidad de un sesgo de selección. Varios factores pueden causar que en un estudio se tengan diferentes “tiempos cero” o “de inicio” tales como: ser un centro de referencia (sólo acuden los más graves o los más complicados), recibir pacientes preseleccionados (sólo atender los casos interesantes), ser el único centro de la localidad (recibir a los afectados por condiciones particulares locales) y admitir por factores socioeconómicos (aspectos raciales, lingüísticos, económicos, de prestación laboral, etc.).

Otro sesgo muy común en los estudios de pronóstico, como se comentó previamente, es la información incompleta de los sujetos incluidos en los estudios. Idealmente, en todo estudio de pronóstico se deben conocer los desenlaces de todos los participantes. Cuando existen pérdidas mayores al 20% de los incluidos, los resultados pueden no ser confiables dado que los sujetos que abandonan o se pierden durante el seguimiento suelen ser aquellos con eventos adversos. Asimismo, el análisis de sobrevida antes comentado asume que los factores de pronóstico estudiados suelen actuar igual durante todo el tiempo de seguimiento, condición no siempre real; por tanto, las pérdidas en diferentes momentos pueden ser explicadas por factores diferentes.

En cuanto a la predicción pronóstica, es importante establecer desde un inicio: ¿cuál(es) es (son) el (los) resultado(s) que se desea predecir o pronosticar? Un grupo de pacientes con una enfermedad puede tener una buena sobrevida (pronóstico favorable), pero con una mala calidad de vida funcional (pronóstico desfavorable); por ello, es fundamental que en un estudio se definan adecuadamente y se midan de manera objetiva los resultados de pronóstico que se están evaluando. Aunque existen múltiples resultados predecibles, éstos se pueden agrupar en cinco clases: sobrevida o defunción, complicaciones específicas, secuelas, consecuencias en la calidad de vida e impactos en la economía.

## Control de sesgos

Con el propósito de reducir los posibles sesgos antes comentados, se dispone de estrategias en el diseño de los estudios. Sin embargo, a pesar de éstas, dada la frecuente existencia de cohortes con un reducido número de pacientes, se han diseñado estrategias metodológicas para controlar el efecto de factores confusores. Las más empleadas son:

1. *Pareamiento de sujetos.* Consiste en disponer de uno o más sujetos con la misma o mismas característica(s) necesaria(s) de controlar y así observar sólo diferencias con respecto a los factores a evaluar.
2. *Restricción de los sujetos.* Consiste en estudiar a individuos que compartan un mismo tiempo cero de su evolución. Con ello, el pronóstico puede ser mejor estudiado al evitar factores externos asociados a las complicaciones previas al comienzo de la cohorte.
3. *Estratificación por factores confusores.* Este método puede usarse al inicio o al final del estudio; consiste en agrupar a los sujetos con un mismo factor pronóstico importante y potencialmente confusor para la evaluación de otro en estudio. Los pacientes



pueden ser agrupados al inicio según el estadio de enfermedad o la presencia de una complicación. Lo anterior permite analizar el pronóstico de pacientes en diferentes “tiempos cero”.

4. *Análisis multivariable*. Esta estrategia se realiza al final del estudio a través del uso de programas estadísticos. El objetivo es aislar el efecto del factor pronóstico de interés considerando los otros factores potencialmente confusores que pudieran afectar su interpretación con respecto al pronóstico evaluado. Existen múltiples métodos disponibles; éstos en general dependen sobre todo de cómo se mide la variable resultado (pronóstico). Cuando la variable resultado se mide de forma dicotómica (sobrevida contra muerte) en tiempos fijos o definidos, se utiliza la regresión logística; cuando interesa saber el tiempo al evento en esta medición dicotómica (tiempo a supervivencia), se dispone de la regresión de Cox; cuando se tiene un resultado de frecuencia de un evento (el cual es poco frecuente), está la regresión de Poisson, y en resultados continuos (días de hospitalización) se puede disponer de la regresión lineal múltiple.

## Conclusiones

Los estudios de pronóstico en *epidemiología clínica* son aquellos que aportan el máximo de información para documentar los factores que pueden asociarse con una probabilidad de desarrollar una enfermedad o fallecer por alguna causa. Para el adecuado análisis de la información es requisito indispensable tener un periodo de seguimiento definido en el tiempo, así como contar con los datos de tiempos de seguimiento de cada paciente en la cohorte hasta el momento de presentar el evento (falla) o, al menos, documentar el momento de la pérdida al seguimiento.

Hay varios recursos de análisis de la información que aportan elementos para estimar los riesgos de incidencia o mortalidad en las cohortes. Por una parte, se tiene la estimación de la incidencia o mortalidad acumulada (*cumulative incidence* o *mortality*) en el tiempo y la comparación de éstas entre dos cohortes (con y sin factor de pronóstico), medida conocida como **riesgo relativo**. Por otro lado, se cuenta con la posibilidad de calcular densidades de incidencia o mortalidad (*hazard rate*) y la comparación entre cohortes (con y sin factor de pronóstico), medida conocida como **razón de densidades de incidencia o mortalidad** (*hazard ratio*). Finalmente, un recurso de análisis estadístico, ampliamente aplicado en este tipo de estudios, consiste en la estimación de la función de supervivencia acumulada a un tiempo  $t$ . El método de Kaplan-Meier es el más conocido. Dicho método permite estimar y graficar la experiencia de supervivencia de una o más cohortes a través del tiempo. Entre sus principales ventajas se encuentra la facilidad de cálculo y de que todos los sujetos de la(s) cohorte(s) aportan información para el cálculo de la medida.

Se debe tener presente la posibilidad de sesgos específicos derivados de los estudios de pronóstico. Hay, sin embargo, estrategias de control en el diseño y en el momento del análisis estadístico que permiten reducir su efecto sobre los estimadores (p. ej., riesgo relativo) y lograr conclusiones válidas.

## Referencias

1. Tan EM, Cohen AS, Fries JF, Massi AT, McShane DJ, Rothfield NF *et al.* The 1982 Revised Criteria for the Classification of Systemic Lupus Erythematosus. *Arthritis Rheum* 25:1271-1277. 1982.
2. Hochberg MC. Updating the American College of Rheumatology Revised Criteria for the Classification of Systemic Lupus Erythematosus [letter]. *Arthritis Rheum* 40:1725. 1997.
3. Alarcón-Segovia D, Estañol B, García-Ramos G, Villa AR. Antiphospholipid antibodies and the antiphospholipid syndrome: clinical relevance in neuropsychiatric systemic lupus erythematosus. *Ann N Y Acad Sci* 823:279-288. 1997.
4. Hanly JG. Neuropsychiatric lupus. *Rheum Dis Clin North Am* 31(2):273-298. 2005.
5. Finley-Caulfield A, Wijman CAC. Management of acute ischemic stroke. *Neurol Clin* 26:345-371. 2008.
6. Rothman KJ, Greenland S. *Cohort studies*. Cap. 7, en: Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3a. ed. Lippincott, Williams & Wilkins. Baltimore, EUA, 2008.
7. Breslow NE, Day NE. *The role of cohort studies in cancer epidemiology*. Cap. 1, en: Breslow NE, Day NE. *Statistical methods in cancer research*. Vol II – The design and analysis of cohort studies. International Agency for Research on Cancer. World Health Organization. Lyon, Francia, 1987.
8. Fletcher RH, Fletcher SW. Cap 7. Pronóstico, en Fletcher RH, Fletcher SW. *Epidemiología clínica*. 4a. ed. Lippincott, Williams & Wilkins, Barcelona, España, 2008.
9. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Lifetime Learning Publications. Belmont, EUA, 1982.
10. Alarcón-Segovia D, Alarcón-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR *et al.* Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL Cohort. *Arthritis Rheum* 52(4):1138-1147. 2005.
11. Alarcón-Segovia D, Pérez-Vázquez ME, Villa AR, Drenkard C, Cabiedes J. Preliminary classification criteria for the antiphospholipid syndrome within systemic lupus erythematosus. *Seminars Arthr and Rheum* 21(5):275-286. 1992.
12. Pérez-Vázquez ME, Villa AR, Drenkard C, Cabiedes J, Alarcón-Segovia D. Influence of disease duration, continued follow-up, and further antiphospholipid testing on the frequency and classification category of antiphospholipid syndrome in a cohort of patients with systemic lupus erythematosus. *J Rheumatol* 20(3):437-442. 1993.
13. Drenkard C, Villa AR, Alarcón-Segovia D, Pérez-Vázquez ME. Influence of the antiphospholipid syndrome in the survival of patients with systemic lupus erythematosus. *J Rheumatol* 21(6):1067-1072. 1994.
14. Arce-Salinas CA, Villa AR, Martínez-Rueda JO, Muñoz L, Cardiel MH, Alcocer-Varela J, Alarcón-Segovia D. Factors associated with chronic renal failure in 121 patients with diffuse proliferative lupus nephritis. A case-control study. *LUPUS* 4:197-205. 1995.
15. Villarreal GM, Drenkard C, Villa AR, Slor H, Shafrir S, Bakimer R, Shoenfeld Y, Alarcón-Segovia D. Prevalence of 13 autoantibodies and of the 16/6 and related pathogenic idiotypes

- in 465 patients with systemic lupus erythematosus and their relationship with disease activity. *LUPUS* 6:425-435. 1997.
16. Gómez-Pacheco L, Villa AR, Drenkard C, Cabiedes J, Cabral AR, Alarcón-Segovia D. Serum Anti-beta B<sub>2B</sub>-glycoprotein-I and anticardiolipin antibodies during thrombosis in systemic lupus erythematosus patients. *Am J Med* 106:417-423. 1999.
  17. Alarcón-Segovia D, Pérez-Ruiz A, Villa AR. Long-term prognosis of antiphospholipid syndrome in patients with systemic lupus erythematosus. *J Autoimmun* 15:157-161. 2000.
  18. Hernández-Cruz B, Tapia N, Villa-Romero AR, Reyes E, Cardiel MH. Risk factors associated with mortality in systemic lupus erythematosus. A case-control study in a tertiary care center in Mexico City. *Clin Exp Rheumatol* 19:395-401. 2001.
  19. Alarcón-Segovia D, Drenkard C, Villa AR. Survival of Mexican patients with systemic lupus erythematosus. *Rheumatology (Oxford)* 40(2):228-229. 2001.
  20. Pons-Estel BA, Catoggio LJ, Cardiel MH, Soriano ER, Gentiletti S, Villa AR *et al.* The GLADEL Multinational Latin American Prospective Inception Cohort of 1 214 Patients with Systemic Lupus Erythematosus. Ethnic and Disease Heterogeneity Among Hispanics. *Medicine* 83(1):1-17. 2004.
  21. Mata-Miranda P, Baptista-González H, Villa AR, Méndez-Sánchez N. Epidemiometría, en: Méndez-Sánchez N, Villa AR, Uribe M. *Métodos clínicos y epidemiológicos de investigación médica*. Elsevier, Masson-Doyma, México, 2006.
  22. Villa AR, Pedroza-Granados J. Análisis de sobrevida, en: Méndez-Sánchez N, Villa AR, Uribe M. *Métodos clínicos y epidemiológicos de investigación médica*. Elsevier, Masson-Doyma, México, 2006.
  23. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53:457-481. 1958.
  24. Katz MH. *Multivariable analysis. A practical guide for clinicians*. Cambridge University Press, EUA, 2006.

# Capítulo 14

## Análisis de decisión en la práctica médica

*Niels H. Wachter*  
*Leticia Adriana Valdez González*



### Introducción

El análisis de decisión es una técnica cuantitativa, para contrastar el valor relativo de diferentes alternativas de acción, en especial cuando existen múltiples objetivos y diferentes resultados esperables, que pueden ser conflictivos.<sup>1</sup>

Uno de los problemas más grandes que enfrenta el médico es la necesidad de tomar decisiones con base en información imperfecta o incompleta. No hay una sola enfermedad que se presente siempre igual con todos sus signos y sus síntomas y el conocimiento sobre las enfermedades aún es incompleto.<sup>2</sup> Además, los padecimientos suelen asociarse en un mismo individuo, es posible que un sujeto sufra varias entidades de manera simultánea y cada una de ellas puede contribuir a las molestias y los demás signos y síntomas de la entidad, lo que incluso modifica las manifestaciones de las otras enfermedades.<sup>3</sup> Lo que es peor, todavía no es posible describir en forma completa y adecuada todos los matices propios de una sola enfermedad.<sup>4</sup>

La variabilidad biológica, la imperfección de los instrumentos de medición, el desarrollo incompleto de métodos de estandarización en los distintos laboratorios y otros factores ocasionan que el resultado de las pruebas auxiliares del diagnóstico no siempre indique con absoluta certeza si el paciente tiene o no la enfermedad.<sup>5,6</sup> Los términos **sensibilidad** y **especificidad** señalan que aun con el resultado de una prueba, todavía hay incertidumbre.<sup>7</sup> En el capítulo 12 de este libro se habló de valores de predicción que indican que los datos observados en los enfermos son de naturaleza probabilística;<sup>8</sup> es decir, que siempre que se obtiene una respuesta a las preguntas (el resultado de una biometría, una medición de enzimas, o la búsqueda de un antígeno en la sangre) existe cierto grado de incertidumbre. Además, hay una gran cantidad de pruebas que informan resultados en una escala continua y no en forma dicotómica (normal, anormal) y en ellas la probabilidad de sufrir la enfermedad aumenta con el valor del resultado de la prueba, de manera que para cada intervalo existe una probabilidad de estar enfermo y otra de no estarlo.<sup>9</sup>

Lo mismo ocurre cuando se quiere establecer el pronóstico o cuando se intenta decidir cuál será el mejor tratamiento,<sup>10</sup> pues en ambos casos, la información está sujeta a incertidumbre. Por ejemplo, saber que un fármaco puede curar a 95% de los pacientes no indica a qué paciente habrá de curar (95/100) y a cuáles no (5/100).

El clínico experimentado aprende a usar esta información de manera intuitiva y toma las decisiones que considera adecuadas de acuerdo con lo que espera que será el resultado más probable,<sup>11</sup> por medio de un proceso que Tversky (en 1974 y 1981) y Schmidt (1990) han llamado **heurística**,<sup>12-14</sup> este es un proceso cognitivo (opera de manera inconsciente) que permite hacer estimaciones de probabilidad con base en la experiencia previa.

No significa que el clínico ejecute un proceso lineal y lógico cada vez que enfrenta un problema diagnóstico; más bien sigue un procedimiento personal que lo lleva a tomar decisiones importantes, aun cuando la información es incompleta o los resultados son inciertos.<sup>15</sup> Este proceso presenta problemas que se discutirán a continuación:

- a) *Heurística de representatividad.* El médico estima si un paciente con el padecimiento "X" tiene en realidad la enfermedad "A". Para esto, contrasta las manifestaciones de tal padecimiento con el prototipo de la enfermedad "A". Si el parecido es importante, realiza el diagnóstico; si no, busca otro. Con esta estrategia ocurren los siguientes errores:
  - Las características de los pacientes suelen estar determinadas por el tipo de práctica clínica (medicina general o de especialidad).
  - Se ignora la probabilidad *a priori* de la enfermedad.
  - Puede compararse el padecimiento con una experiencia pequeña y poco representativa.
  - No permite incorporar las modificaciones de cuadro clínico que son consecuencia de la comorbilidad.
- b) *Heurística de memoria.* La experiencia se refiere a los hechos que ocurrieron en el pasado y es más fácil evocar los casos poco usuales y los que ocurrieron en circunstancias especiales, mientras que los más comunes quedan relegados en el olvido.
- c) *Heurística de ancla y ajuste.* La historia clínica genera mucha información y el clínico usualmente selecciona los síntomas más sobresalientes del caso para estimar cuáles podrían ser los posibles diagnósticos. Luego ajusta esta probabilidad con las características particulares del caso (p. ej., casi todos los enfermos con síndrome febril suelen tener, además, cefalea, mialgias y artralgias). Con esta estrategia puede ocurrir lo siguiente:
  - Elegir mal el "ancla" (p. ej., seleccionar la cefalea y no la fiebre como la característica sobresaliente).
  - No ajustar la probabilidad con las características particulares del caso (la fiebre suele tener muchas causas, que dependen de las características particulares del

caso; por ejemplo, si se trata de una persona joven o de mayor edad o si se trata de alguien del medio urbano o rural).

La experiencia no es la única fuente de información; también se usa la información publicada en revistas y libros de texto, que también tiene limitaciones.<sup>16</sup> De esta manera la mayoría de los casos se puede resolver más o menos con rapidez. El procedimiento se ve reforzado por la experiencia y por la repetición y, en muchas ocasiones, el resultado se consigue de manera casi instantánea al conocer la información más sobresaliente durante la entrevista con el enfermo.

Existe una proporción sustancial de casos en los que la experiencia previa no ayuda, como los que se mencionan en el recuadro siguiente:

- Porque no se parecen en nada a lo visto antes.
- Los resultados difieren en distintos subgrupos de pacientes.
- El padecimiento se complica con otras enfermedades o circunstancias.
- Las consecuencias del tratamiento (beneficiosas y perjudiciales) difieren en diferentes pacientes para cada tratamiento; de manera que se recomienda en unos casos y no en otros.<sup>17</sup>
- O bien simplemente porque existe tanta información que ya no es posible procesarla en forma sistemática en el pensamiento.<sup>18-20</sup>

Tales son los casos difíciles en particular, en los cuales resulta útil una herramienta que se desarrolló en la investigación de operaciones y la teoría de los juegos. Dicho instrumento se llama **análisis de decisión** y se vale del mismo procedimiento empleado para comprender la utilidad de las pruebas de diagnóstico, el teorema de Bayes o alguna otra técnica para estimar probabilidades. Con frecuencia utiliza representaciones gráficas (árboles de decisión) y se ha visto impulsada por la disponibilidad de computadoras personales, así como por el desarrollo de técnicas de inteligencia artificial que han facilitado en gran medida los cálculos necesarios para efectuar el análisis.<sup>21</sup>

Los estudios epidemiológicos ofrecen información sobre factores de riesgo en una población; la mayoría de los estudios clínicos informa factores pronósticos y resultados de tratamiento en condiciones específicas, con el resultado de que la población se convierte en un conjunto heterogéneo de sujetos con diferentes factores de riesgo y enfermos con diferente pronóstico y posibles resultados de tratamiento. ¿Cómo decidir qué hacer en cada caso? En términos generales, cuando los beneficios superan a los efectos adversos se prefiere una estrategia que favorezca la sensibilidad, es decir, que todos reciban el tratamiento (p. ej., en la apendicitis aguda) mientras que, cuando los efectos adversos son considerables, se prefiere una estrategia que favorezca la especificidad; es decir, se elige con mucho cuidado quién deberá exponerse al tratamiento (p. ej., la quimioterapia del cáncer).

## Caso clínico

Considere el caso de un varón de 58 años, quien hace dos meses sufrió infarto al miocardio no complicado y desde entonces tiene angina de esfuerzo; no presenta ninguna otra anormalidad cardiovascular. Hace siete semanas inició el dolor lumbar que ha progresado con irradiación a la pierna derecha por debajo del hueco poplíteo; tiene anestesia del primer orfejo y de la cara lateral de la pierna. Además se nota debilidad en la flexión del dorso del pie. Su médico le indicó algunos días de reposo con analgésicos antiinflamatorios. Esta molestia dificulta el programa de rehabilitación del enfermo y, a pesar de todos estos esfuerzos, el individuo continúa con dolor y déficit neurológico. Se ordena imagen de resonancia magnética nuclear que demuestra hernia de disco en el segmento L4-L5. Un estudio electromiográfico del segmento inferior es compatible con compresión radicular del segmento afectado.

La conjunción de síntomas, imagen de resonancia magnética y electromiografía, así como la progresión de la enfermedad y la falla del tratamiento conservador, hacen suponer que se trata de un caso con indicaciones de operarse; sin embargo, se sabe que el infarto al miocardio reciente implica un riesgo elevado y no se reconoce cuál es la mejor decisión: si operar o no al enfermo. Es este momento se está en capacidad de formular un problema con mucha precisión y de identificar las alternativas de acción (primer paso del análisis).

## Análisis de decisión

Esta técnica se aplicó por primera vez para examinar y determinar cuándo hacer cirugía radical en pacientes con cáncer de la boca, sin metástasis palpables en el cuello, y a partir de entonces, se ha aplicado cada vez con más frecuencia en el campo de la Medicina Clínica porque tiene algunos atractivos muy interesantes;<sup>22-26</sup> obliga a identificar y delinear con claridad el problema en el tiempo y, a veces, a buscar información adicional que inicialmente no se había solicitado así como a mostrar de manera explícita las alternativas de acción y todas sus posibles consecuencias, a estimar la probabilidad con la que cada una de estas consecuencias ha de ocurrir y, sobre todo, a tomar una decisión con base en un procedimiento lógico cuantitativo. Además permite variar las condiciones del análisis en forma teórica para conocer cuál sería el resultado si tales condiciones fueran distintas (esto se conoce como *análisis de sensibilidad*).<sup>27</sup>

El análisis de decisión está indicado cuando se conocen los componentes originales de la decisión (es decir, las opciones relevantes de tratamiento, las consecuencias de cada una de ellas y la probabilidad con la que podrían ocurrir).<sup>28</sup>

A pesar de todas estas bondades, aún existen problemas no resueltos que se constituyen en desventajas de la técnica.

- La técnica es sólo una herramienta, el médico retiene toda su responsabilidad por las decisiones que toma en cada caso.<sup>29</sup>

(continúa)



- Los cálculos necesarios pueden consumir mucho tiempo y no todos los clínicos disponen de los conocimientos ni del *software* especializado para hacerlo.<sup>30</sup>
- El lenguaje y las técnicas utilizados pueden resultar difíciles de comprender para el clínico, así como los modelos que se presentan.
- El clínico se siente intranquilo al usar información que se obtiene de publicaciones y se procesa en forma teórica.
- Todo análisis supone simplificar un problema complejo; si el clínico decide eliminar de este análisis alternativas de acción o consecuencias que eran importantes, los resultados serán irrelevantes.
- El resultado final del análisis depende fuertemente de cómo se determinen las utilidades y éstas pueden variar de acuerdo con el método usado para obtenerlas, de un país a otro, entre grupos de edad y situación social específica.

A pesar de todo esto, la técnica continúa desarrollándose, aun cuando el solo uso de los números no le otorga legitimidad.<sup>27</sup>

Con estas técnicas se han desarrollado reglas de decisión clínica de gran valor,<sup>1</sup> que permiten actuar con más seguridad en condiciones de incertidumbre y en ocasiones han permitido prescindir de auxiliares del diagnóstico que no aportan información adicional a la que se obtiene de la anamnesis cuidadosa.<sup>31,32</sup>

Sin embargo, las reglas de predicción y las reglas de decisión deben someterse a un proceso de validación que demuestre su utilidad.<sup>33</sup> Se ha demostrado que el médico consume gran cantidad de información y que además la usa con poca frecuencia en los casos complejos.<sup>34</sup> El clínico que utiliza esta información de acuerdo con una regla de decisión clara y explícita obtiene resultados más reproducibles y certeros que quien se basa en un procedimiento intuitivo que suele arrojar para el mismo caso resultados en ocasiones diferentes.<sup>35</sup>

El análisis de decisión permite cuantificar la incertidumbre, de manera que es más fácil comunicar resultados a los demás, que con el solo uso de términos como **posible**, **probable** o **frecuente**.

El análisis explícito y cuantitativo obliga a identificar las características sobresalientes del caso, así como a reflexionar sobre las alternativas de tratamiento y sus consecuencias; de este modo, siempre y cuando se usen información certera y un modelo adecuado, es posible obtener conclusiones válidas y generalizables.

El análisis de decisión se fundamenta en maximizar la utilidad esperada de una alternativa de acción (aplicar o no una prueba diagnóstica, elegir uno u otro tratamiento), que puede expresarse de diferentes maneras: en Economía la utilidad representa ganancia o pérdida monetaria; en Psicología representaría satisfacción u otro constructo muy parecido; mientras que en Medicina, usualmente se representa como mortalidad, morbilidad, sobrevida, años de vida ganados o perdidos o la proporción de casos evitados o curados.



## Técnicas del análisis de decisión

Existen diferentes técnicas para llevar al cabo un análisis de decisión, pero las más usadas en medicina son el árbol de decisión (Bayes) y los procesos de Markov.

Antes de iniciar el análisis, deben resolverse las interrogantes mencionadas en el recuadro siguiente:

- Delimitar el problema con precisión. Esta es la parte inicial del análisis, ha de ser claro y explícito, y debe incluir toda la información necesaria para distinguirlo de otros similares, pero que pueden tener soluciones o consecuencias diferentes.
- Identificar las alternativas de acción y todas sus consecuencias relevantes. Constituye la parte medular del análisis. Para que éste se lleve a cabo, debe existir un conflicto, un punto en el que no se sabe cuál de dos o más alternativas ha de escogerse (una de las alternativas puede ser no hacer nada y esperar el resultado). Entre las consecuencias que pueden estudiarse están los efectos mismos del tratamiento (curación, mejoría parcial, falla, muerte, efectos adversos) y los costos de cada alternativa. Este último punto se discute en el capítulo 5.
- Identificar el punto de vista del análisis: se hará desde la perspectiva del paciente, del sistema de salud que paga la atención o desde el punto de vista de la sociedad.
- ¿De dónde habrá de obtenerse la información?: algunos estudios se hacen como complemento de un ensayo clínico controlado y en otros deberá buscarse la información publicada. En este último caso, deben considerarse: estrategias de búsqueda, palabras clave, la calidad que se espera de las fuentes originales de información, etc., de la misma manera que se haría para un estudio de metaanálisis.
- ¿Con qué técnica se hará el análisis?

## Para construir un árbol de decisión

Un árbol de decisión es la técnica más sencilla de análisis de decisión; es una representación gráfica que permite visualizar directamente el problema, las alternativas de acción y las consecuencias. Para la construcción de este árbol se usa una serie de símbolos (nodos) y líneas de conexión entre ellos, que tradicionalmente se representan con líneas rectas. Esta técnica suele usarse cuando se visualiza el problema de manera análoga a un estudio transversal.

A fin de efectuar el análisis de decisión deben seguirse los pasos que muestra la figura 14-1.<sup>2,21,27</sup>

En cuanto a los incisos *a* y *b*, el problema principal y sus alternativas de acción se representan con figuras cuadradas o rectangulares. Se llaman **nodos de decisión** (representan las alternativas de acción en los procesos que están sujetos a la voluntad de quien efectúa el análisis). En un árbol, éstas serían el tronco y las ramas principales.

Sobre el inciso *c*, las consecuencias de cada alternativa de acción se representan con figuras circulares que se denominan **nodos de probabilidad**, dichas figuras representan

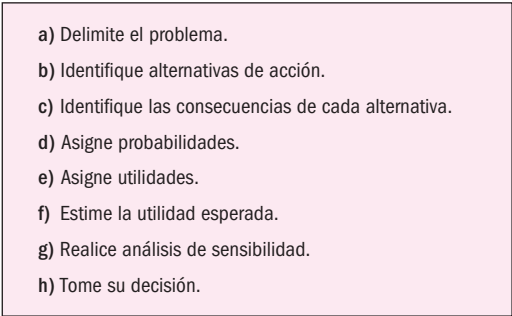
- 
- a) Delimite el problema.
  - b) Identifique alternativas de acción.
  - c) Identifique las consecuencias de cada alternativa.
  - d) Asigne probabilidades.
  - e) Asigne utilidades.
  - f) Estime la utilidad esperada.
  - g) Realice análisis de sensibilidad.
  - h) Tome su decisión.

**Figura 14-1.** Pasos a seguir para efectuar el análisis de decisión.

las “ramas” del árbol. Cuando el problema es complejo o las consecuencias se repiten dentro de cada una de las ramas, pueden construirse subunidades repetitivas (“subárboles”). Si la estructura del árbol se hace compleja, la línea que une a este nodo con los que le siguen se “rompe” con una doble diagonal (-//-) y se dibuja el subárbol correspondiente en otra página. El punto final de cada “rama” (por ejemplo, la curación o el fallecimiento) suele representarse con una figura triangular que se denomina **nodo terminal**.

En el inciso *d* se busca **asignar probabilidades**; una vez que se tienen el tronco y las ramas, se deben colocar las “hojas”; esto es, debe estimarse la probabilidad con la que ocurrirá cada una de las consecuencias que se identificaron previamente.

En este punto deben considerarse los siguientes factores:

- La información debe ser de alta calidad; es decir, deben consultarse fuentes originales (artículos de diagnóstico, pronóstico o tratamiento) con información válida y precisa.
- Cuando no existe información publicada, se pueden consultar las bases de datos de las instituciones de salud, de las compañías de seguros médicos o estadísticas vitales y, en última instancia, de expertos del tema que generen (de acuerdo con su conocimiento previo del campo) esas probabilidades.<sup>29</sup>
- Deben identificarse fuentes de incertidumbre. Cuando hay más de una referencia o cuando la información se generó a partir de la opinión de un grupo de expertos, es difícil que los resultados de uno y otro concuerden exactamente. Aun en el caso de que sólo hubiera una referencia útil, los datos se obtienen de muestras que están sujetas a variabilidad e incertidumbre y quien hace el análisis deberá identificar entre qué valores puede variar esta cifra (p. ej., los límites del intervalo de confianza al 95%).

Todas las probabilidades que parten de un solo nodo deben sumar la unidad (1.0 o 100%). Así, si después de aplicar una prueba de diagnóstico, se estima que 90% resultará positivo, el nodo complementario (los que resultan negativos) tendrá una probabilidad de 10%. Así se continúa con cada subnodo; por ejemplo, para el primer nodo “prueba positiva”, si 90% tiene la enfermedad (verdaderos positivos), 10% no tendría la enfermedad (falsos positivos).

El inciso *e*, **asignar “utilidades”**, implica que no todas las consecuencias de la enfermedad o del tratamiento tienen el mismo valor para los pacientes. Cuando les asignamos un valor, incorporamos las preferencias de las personas al comparar los méritos relativos de diferentes alternativas de acción. Parecería relativamente sencillo asignar valor a una curación sin secuelas (1.0) y a la muerte (0.0). El problema reside en asignarles valor a todos los posibles resultados intermedios, que son muy comunes en las enfermedades crónicas. Una sección subsecuente describirá los métodos para estimar utilidades.

En cuanto al inciso *f*, estimar la “**utilidad esperada**”, para esto es necesario calcular la probabilidad con que ocurrirá cada proceso en toda rama del árbol. La literatura en inglés denomina a este procedimiento *fold back* (“doblar hacia atrás”) y consiste en multiplicar de derecha a izquierda la probabilidad de cada nodo por el previo hasta llegar al nodo de decisión. El resultado se anota hacia la extrema derecha de cada rama. Los valores obtenidos deben sumarse 1.0 en cada nodo de decisión.

Cada uno de estos valores se multiplica por la utilidad asignada a tal consecuencia del tratamiento y se suma dentro de cada nodo de decisión. Al valor resultante se le llama **utilidad esperada** y representa un promedio ponderado de la utilidad que se obtendría si se tomara tal decisión (la suma de la máxima utilidad que se obtendría al tomar esa decisión).<sup>2,21,27</sup>

En el inciso *g*, efectuar **análisis de sensibilidad**, es posible variar las condiciones del análisis y es posible modificar la sobrevida o efectividad del tratamiento, o bien los valores de utilidad (se pueden usar los límites del intervalo de confianza de 95% de los datos que se encontraron o los límites con que los expertos indicaron que podría ocurrir tal desenlace). Se repite todo el procedimiento con esta nueva información a partir del inciso *d*.

Si una de las alternativas continúa siendo superior a las demás, ya se tiene una estimación estable. Si cambia el orden en que aparecen las preferencias (la alternativa con la utilidad esperada más alta), se tiene una situación que depende de la probabilidad con la que ocurran estos procesos. En este caso deben conocerse el punto exacto en que la preferencia cambia por una u otra alternativa de acuerdo con la probabilidad con que suceden las consecuencias o la utilidad que éstas tienen (el punto en que cambia la preferencia se llama “umbral”).<sup>36</sup>

Por último, en el inciso *h*, se debe **tomar la decisión** de comparar los valores de utilidad esperada para cada nodo de decisión y decidirse por el más alto.

## Análisis de Markov

Un modelo de Markov es aquel en el que se asume que un número de pacientes se encuentra dentro de un número finito de posibles estados de salud (estados de Markov), que deben ser exhaustivos (todos los posibles) y mutuamente excluyentes (es decir, no se puede pertenecer a dos de ellos simultáneamente). Estos estados pueden ser “absorbedores”; es decir, una vez que se alcanzan, no se puede regresar, o “no absorbedores”, pues se puede entrar y salir de ellos. Los acontecimientos se representan como pasos (“transiciones”) en periodos uniformes (ciclos de Markov) que ocurren con una probabilidad dada

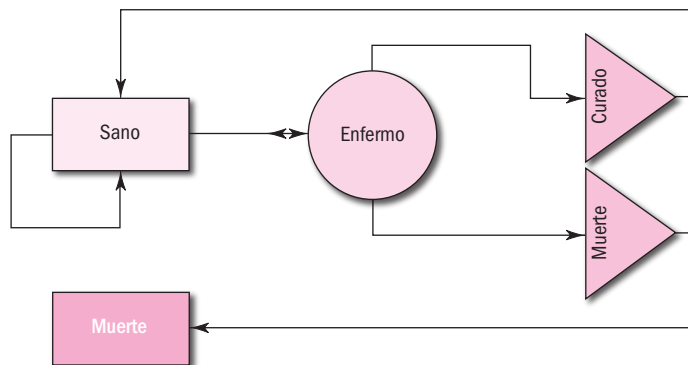


**Figura 14-2.** Representación esquemática de procesos de Markov. El análisis parte del sujeto sano que puede hacer una transición a enfermar. Una vez ahí, puede regresar a sano o fallecer. Este último estado es “absorbedor”, pues no es posible regresar; dicho proceso se repite en cada ciclo y el individuo puede permanecer sano o enfermar, mientras que los casos de muerte se van acumulando.

para cada ciclo. Se pueden representar gráficamente como una variante de los árboles de decisión (figura 14-2).

A diferencia del árbol de decisión que se definió como el análogo de un estudio transversal, en este caso podría visualizarse como un estudio longitudinal. Por ejemplo, una persona sana puede permanecer así o ingresar al estado de enfermedad y al curarse podría regresar al estado de sano o podría trasladarse al estado de muerte, del cual no habría regreso. Estas transiciones se repiten en cada ciclo, con una probabilidad conocida, por varios ciclos (tantos como se hubieran planeado en el análisis original) hasta el final del análisis.<sup>37</sup> Otra diferencia importante entre el análisis de decisión y el análisis de Markov reside en el hecho de que mientras el análisis de decisión es una herramienta de “optimización”, es decir, siempre resulta en una recomendación, el análisis de Markov es una técnica “descriptiva”, pues sólo provee la información probabilística y no asume una “recomendación” final.

Condiciones para el análisis de Markov:<sup>38</sup>

- a) Está justificado cuando se trata de eventos repetitivos, irreversibles y de larga duración.
- b) Sólo se permiten transiciones entre estados previamente establecidos.
- c) La duración de los ciclos es arbitraria, pero debe ser constante en todo el análisis.
- d) Cada paciente sólo puede hacer una transición en un ciclo.
- e) Todos los pacientes están sometidos a las mismas probabilidades de transición.

Existen dos tipos de análisis de Markov: el análisis se llama **cadenas de Markov** cuando las probabilidades de transición permanecen constantes en todos los ciclos y **procesos de Markov** cuando estas probabilidades cambian con el tiempo (por ejemplo, la tasa de letalidad de una enfermedad puede aumentar conforme el paciente se hace más viejo).

Para hacer un análisis de Markov se siguen estos pasos:

- a) Definir los estados de salud.
- b) Establecer las transiciones permitidas entre cada estado.
- c) Definir la duración de los ciclos (horas, días, meses, años) y su duración (horizonte temporal) final. Esto de acuerdo con la historia natural de la enfermedad y la probabilidad real de sobrevida de los sujetos en estudio.
- d) Identificar las probabilidades de transición entre los diferentes estados.
- e) Asignar utilidades y, cuando corresponda, también los costos.
- f) Resolver el modelo.
- g) Hacer el análisis de sensibilidad.
- h) Comparar resultados y tomar una decisión.

El modelo se puede resolver con cualesquiera de tres estrategias:

- a) La simulación de una cohorte hipotética.<sup>39</sup>
- b) Simulación de Montecarlo.<sup>40</sup>
- c) Álgebra de matrices (el proceso original descrito por Andrei Markov).

Estos cálculos pudieran resultar excesivamente complejos para el clínico y en la práctica, se usa *software* especializado como TreeAge Pro u otros.<sup>41</sup> Para quien esté interesado, también puede hacerse con un paquete Excell (<http://www.cs.brown.edu/research/ai/pomdp/tutorial/>).

## Análisis de sensibilidad

No importa qué técnica se use para el análisis (árboles de decisión, análisis de Markov, redes neurales); en todos los casos se construyó un modelo con base en ciertos supuestos (estados de salud, transiciones, probabilidades) y la información se obtuvo de la literatura y está sujeta a incertidumbre o variabilidad que puede tener su origen en tres fuentes principales.<sup>42,43</sup>

- a) los parámetros,
- b) metodológica,
- c) estructural.

La primera fuente de incertidumbre, los parámetros, tiene que ver con las cifras que se eligieron como apropiadas o representativas para el modelo (las probabilidades que se incluyeron en el modelo). La segunda, metodológica, tiene que ver con los supuestos que fundamentan el modelo; es decir, qué alternativas de acción y qué consecuencias habrán de tomarse en cuenta, y la última, estructural, tiene que ver con la distribución de los datos. En el capítulo 10 se habló de la distribución normal (Gauss) y otras (Poisson y binomial). Sin embargo, cuando se trata de eventos que ocurren con una probabilidad que depende del tiempo transcurrido; como la mortalidad (donde la tasa crece a medida que la persona envejece), lo más probable es que éstos no sigan una distribución normal y se comporten

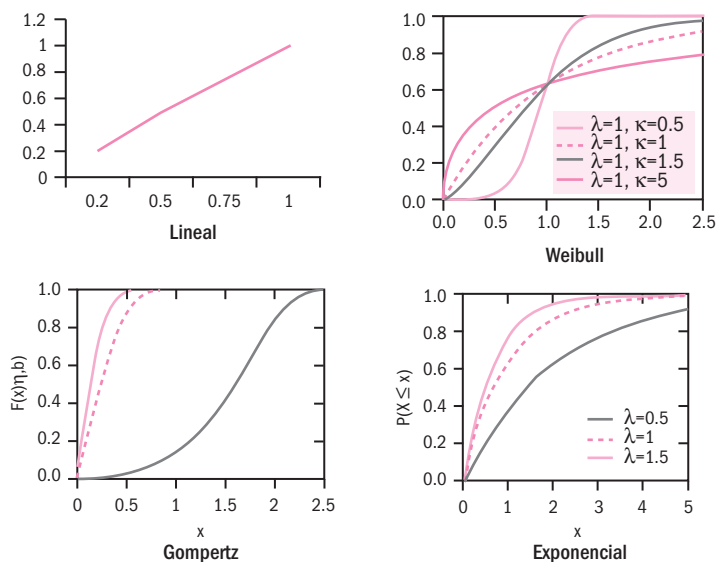


**Figura 14-3.** Diferentes distribuciones de probabilidad.

como una distribución exponencial, Weibull o Gompertz (figura 14-3). Si el analista elige una distribución inapropiada, su estimación del evento final (por ejemplo, la mortalidad o la incidencia de complicaciones) será inexacta y las conclusiones del análisis serán incorrectas.<sup>44</sup>

Para estimar el efecto de la incertidumbre en el resultado final se hace un análisis de sensibilidad.

### Análisis de sensibilidad de una vía

Suponga que la publicación que se consultó señala que la mortalidad era 0.08 y los límites del intervalo de confianza al 95% de esa estimación abarcan desde 0.06 hasta 0.1; en un caso así, sabemos que con 95% de probabilidad, la mortalidad podría ser tan baja como 6% o tan alta como 10%. Si se repite este análisis dos veces, tomando como estimado de mortalidad 6 y 10%, habría hecho un análisis de sensibilidad de una vía. Éste se puede repetir de manera individual para cada una de las variables que se considere que podrían determinar el resultado final. En cada caso se verá si las conclusiones del estudio cambian.<sup>39</sup>

Si ninguno de esos análisis adicionales muestra que cambien las conclusiones (por ejemplo, que la estrategia A siempre es mejor que la estrategia B), se dice que se tiene un resultado “robusto” y puede hacerse la recomendación correspondiente (figura 14-4).

Si una de las variables que cambian sus valores en el análisis de sensibilidad determina una conclusión distinta (p. ej., que la estrategia B sea superior a la A), es necesario hacer un análisis de gatillo. Por ejemplo, cuando se aplica una prueba de escrutinio,

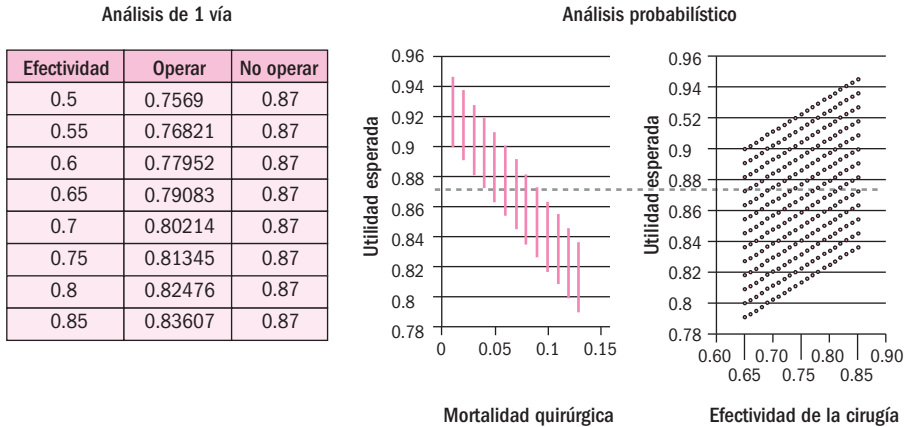


**Figura 14-4.** Análisis de sensibilidad. Para el análisis de una vía, se repitió el análisis con diferentes valores para la efectividad de la cirugía y se calcularon utilidades esperadas para cada una de las opciones de tratamiento. Para el análisis probabilístico se calcularon los valores de utilidad esperada para diferentes valores de efectividad de la cirugía y mortalidad quirúrgica, con una ecuación de regresión. Después se hizo una simulación de Montecarlo, variando al mismo tiempo y de manera aleatoria los valores de efectividad de la cirugía entre 0.5 y 0.85 y los de la mortalidad quirúrgica entre 0.005 y 0.13. Este proceso se repitió en 1000 copias y para cada una de estas copias se calculó la utilidad esperada con la ecuación de regresión generada previamente. Note que en diferentes condiciones la utilidad esperada excede el valor de 0.87 que se obtiene con la opción de no operar (línea punteada). En este caso, habrá que hacer un análisis de gatillo para encontrar a partir de qué valor (especialmente de la mortalidad quirúrgica) es preferible la opción de operar a la de no operar.

ésta suele ser apropiada cuando la prevalencia de la enfermedad es suficientemente alta y cuando la prueba es lo bastante sensible. Cuando la prevalencia o la sensibilidad no son suficientemente altas, es preferible no usar la prueba de escrutinio, pues la probabilidad de resultados falsos positivos o negativos sería excesiva. Un análisis de gatillo permitiría encontrar a partir de qué prevalencia o sensibilidad es útil la prueba de escrutinio (figura 14-5).

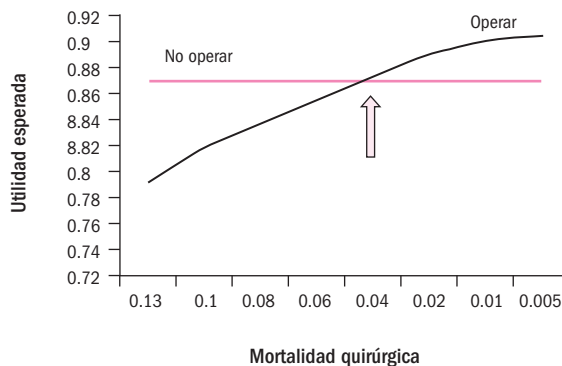


**Figura 14-5.** Análisis de gatillo.

## Análisis de sensibilidad probabilístico<sup>45</sup>

También se puede evaluar cuál sería el efecto de variar al mismo tiempo y de manera aleatoria cada una de las posibles determinantes del resultado y así examinar qué efectos tienen en el resultado final. Para esto sería necesario usar técnicas de regresión múltiple o de simulación de Montecarlo (p. ej., esto puede hacerse en una hoja Excel con la función ALEATORIO.ENTRE), donde se considera qué efecto tienen en el resultado final, cambios al azar, dentro de límites preestablecidos, de cada una de esas variables de manera simultánea (figura 14-5).<sup>46</sup>

## Utilidades

Una de las ventajas de las técnicas de análisis de decisión es que pueden incorporar los valores y preferencias de las personas en el proceso de decisión. Un estudio original mostró que al incorporar las preferencias del paciente, la decisión de usar anticoagulantes en pacientes con fibrilación auricular difiere mucho de la que resultaría de usar *Guías de Práctica Clínica* desarrolladas sólo por profesionales de la Medicina.<sup>47</sup> Las personas valoran de diferente manera sus necesidades y con frecuencia los médicos usamos un criterio diferente.<sup>48</sup> Para los médicos ha resultado todo un reto incorporar las preferencias de los pacientes en las decisiones de tratamiento.<sup>49</sup>

Las preferencias del paciente se integran al estudio al calcular Utilidades. Las utilidades se describen como el valor que una persona le asigna a un determinado estado de salud y como ésta es una percepción subjetiva, es necesario usar técnicas específicas, a fin de obtener resultados válidos y creíbles.

Antes de decidirse por una técnica específica, debe ponerse atención a los siguientes aspectos:

- a) *El estímulo.* Usualmente el médico pide a una persona que imagine que se encuentra en un estado de salud específico y que evalúe sus respuestas a partir de esta información. Cómo se describe este estado y quién debe juzgarlo, determinan la magnitud de la respuesta. Los enfermos que tienen la experiencia porque han vivido en esa condición a menudo asignan utilidades (valores) más altas que cuando se solicita a una persona sana que imagine una situación hipotética de enfermedad (es más fácil juzgar una realidad que imaginar lo que pudiera ocurrir en una situación hipotética). Cómo se presenta la situación (aspectos negativos y positivos, amplitud de la descripción, contexto) también determina la naturaleza de la respuesta. Debe considerarse que las personas juzgan más el efecto de la enfermedad en su vida, que en su salud en general, como haríamos los médicos.
- b) *Interpretación e integración de la información.* Sólo a los investigadores les interesa evaluar utilidades y los pacientes son de ayuda en este sentido cuando responden cuestionarios; esto no es una necesidad del paciente. Así que al tener que resolver el problema mental de ofrecer una respuesta, en el tiempo limitado de la consulta, los pacientes suelen valerse de mecanismos inconscientes que pueden distorsionar la realidad. Uno de ellos es la heurística, misma que se describió antes en este capí-



- tulo. Además, la capacidad de una persona para adaptarse y adoptar una situación teórica para ofrecer una respuesta depende de varios factores personales (estrés, mecanismos adaptativos).
- c) *Juicio*. Las personas tienen que ser capaces de comparar su situación actual con la que se les pide que juzguen y ofrecerán respuestas en función del ánimo que manifiesten cuando se les entrevista.
  - d) *Forma de respuesta*. Los resultados también dependen de cómo se muestran las opciones de respuesta. Algunas personas tienen serias dificultades para comprender algunos formatos (como *time trade-off*). La manera de buscar la respuesta también afecta el resultado, pues si partimos de los valores más altos (100, 90, 80...) los pacientes asignan utilidades más altas que si se muestran opciones altas y bajas de manera alternativa (100, 50, 90, 60, etc.).
  - e) Aunque en apariencia los valores que se obtienen son cuantitativos, lo que realmente obtenemos es en qué orden se encuentran las preferencias de las personas. Casi siempre se consideran dos extremos, completamente sano (que suele tener un valor de 1.0) y la muerte (que suele tener un valor de cero). Si una persona señala que estaría dispuesta a dar la mitad de los años que le queden de vida para “curarse” de una enfermedad, esto no significa que, en efecto, ella pueda intercambiar años de vida por un resultado específico. Este valor se interpreta como que esta persona valora un año de vida en esta situación, con un valor aproximado de 50% de un año con perfecta salud. Este tipo de datos se usan para estimar los llamados QALY (*Quality Adjusted Life Years*) o AVAC (años de vida ajustados por calidad).

Entre las técnicas de evaluación de utilidades están las mencionadas en la figura 14-6.

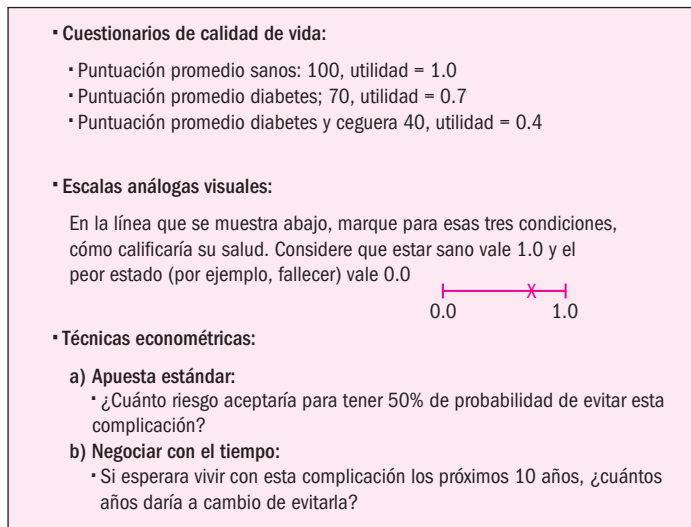


**Figura 14-6.** Técnicas para estimar utilidades.

- a) *Cuestionarios de calidad de vida*. Son cuestionarios previamente validados que evalúan el desempeño de las personas en diferentes áreas de la vida diaria (sensación de bienestar, movilidad, independencia, capacidad para llevar a cabo rol social o familiar, etc.). Existen las dirigidas a uso general, como el SIP (*Sickness Impact Profile*) o específicas de una enfermedad. Para este tipo de análisis se considera la puntuación promedio de las personas sanas como 100% (o 1.0), y las puntuaciones promedio de las personas en diferentes condiciones se convierten a un porcentaje o proporción de ese total (p. ej., una persona sana arrojaría un valor de 1.0, una persona con diabetes arrojaría valores cercanos a 0.75 y una persona con diabetes y ceguera daría valores cercanos a 0.4).
- b) *Escalas análogas visuales*. Se presentan al paciente viñetas (resúmenes) clínicos y se le pide que responda colocando una marca en una línea de una longitud estándar (digamos 10 cm), entre dos opciones (sano = 1.0, muerte = 0.0). Experimentos de Psicofísica (p. ej., con luz de intensidad variable) han demostrado que la magnitud de la respuesta en centímetros tiene una relación logarítmica con la magnitud del estímulo. Así, si el paciente coloca su respuesta a 4 cm del extremo con valor 0.0, el resultado sería 0.4.
- c) *Técnicas econométricas*.<sup>21,50</sup>
- *Apuesta estándar (standard gamble)*. Después de leer una viñeta clínica, se pide al paciente que señale si estaría dispuesto a someterse a un tratamiento con tal de librarse de esa condición (p. ej., ceguera). La efectividad del supuesto tratamiento se fija en un valor por debajo del 100% (podría ser 50%) y después se le señalan distintas probabilidades de riesgo de ese tratamiento. En términos generales, las personas aceptan mayores riesgos para las condiciones más graves (aceptarían en promedio, riesgos de 60% para ceguera [para una “utilidad” de 0.4] y no aceptarían ningún riesgo para un catarro [para una utilidad de 1.0]). La magnitud del riesgo que cada persona acepta está en función de su propia “aversión al riesgo”; es decir, que para una misma situación, unas personas aceptarían más y otras menos riesgo. Cuando se aplica una técnica de esta naturaleza, el valor que se toma es el llamado “punto de indecisión”; es decir, aquel porcentaje en el que ya no sabría si tomar la decisión o no.
  - *Negociar con el tiempo (time trade-off)*. Esta técnica es análoga a la anterior; pero, en lugar de trabajar con riesgos, se le pide a la persona que imagine que vivirá exactamente 10 años más en la situación que se le muestra (sano, diálisis, ceguera, amputación, etc.), pero que podría vivir sin esa condición si ofrece una parte de su sobrevivida. En general, una persona ofrece más años cuando considera ese estado de salud más grave (en este ejemplo, ofrecería 6 de esos 10 años, a fin de librarse de la ceguera, para una utilidad de 0.4).

Por supuesto, todos prefieren curarse, pero la mayoría de las maniobras terapéuticas actuales son imperfectas. Existen otras técnicas para obtener estos valores, que llamaremos “utilidad”, y casi todas arrojan resultados comparables.<sup>51</sup>

## Ajuste en el tiempo

En el análisis es importante considerar en qué momento ocurren las cosas. La ceguera siempre es una consecuencia grave de una enfermedad, pero cuando le sucede a una persona joven en edad productiva es aún más grave que cuando ocurre en la edad avanzada. Cuando el análisis incorpora costos, el pago de un tratamiento el día de hoy se considera más caro (p. ej., el costo de una cirugía) que cuando se paga la misma cantidad, pero de manera fraccionada a través de los años. En este tipo de análisis se aplican tasas de descuento para eventos que ocurren en el futuro. La tasa más usual es de 5% y en el análisis de sensibilidad suelen examinarse variaciones entre 3 y 7%.

Este descuento no corresponde con el cálculo del efecto que pueda tener la inflación en costos futuros y que debe incorporarse por separado al análisis.

## Algunas notas de precaución

El análisis de decisión depende del modelo que se propuso. Un modelo es una representación teórica de la realidad, y el éxito del análisis de decisión dependerá de diversos factores:

- Que todas las alternativas relevantes de tratamiento y todas sus consecuencias importantes estén consideradas. Si el análisis olvida o ignora algunas de ellas, el resultado puede ser irrelevante, inútil o, en el peor de los casos, puede llevar a una conclusión errónea y a una decisión inadecuada. Si incluyen consecuencias de poca importancia, el árbol es tan complejo que nunca podrían efectuarse todos los cálculos necesarios para obtener el resultado.
- Que los estimados de probabilidad de cada nodo sean suficientemente exactos. Si la información es imprecisa o está sesgada, todo el análisis es inútil. Por esta razón debe aplicarse todo el empeño y buen juicio en la representación del árbol de decisión y tener que hacerse todo el esfuerzo necesario con el fin de obtener información de buena calidad. Aun cuando esto no sea del todo posible, se ha dicho que el análisis explícito continúa siendo mejor.<sup>52-54</sup> Se debe procurar que el análisis de sensibilidad incluya los extremos entre los que razonablemente ocurrirán los hechos de interés más importantes. Si el resultado no cambia (esto es, la utilidad esperada de una alternativa continúa siendo más alta que las demás), las conclusiones pueden utilizarse con confianza; si no es así, hay que buscar más información.<sup>2,27,55</sup>
- En algunas situaciones, el árbol de decisión resulta tan complejo que ya no es aplicable el análisis bayesiano (el tiempo y esfuerzo necesarios para computar los detalles pueden hacerlo prohibitivo). En estos casos, consulte a un experto que posiblemente le recomendará otra técnica (procesos de Markov, inteligencia artificial, etc.).<sup>38,56</sup>
- Recuerde que los valores que usó en el análisis y los que obtiene de éste son promedios de probabilidad; esto es, la probabilidad de que ocurra el hecho en la mayoría de los casos o de acuerdo con las condiciones en las que se obtuvo la información, y de ninguna manera señalan el resultado de un caso en particular. De esta manera,

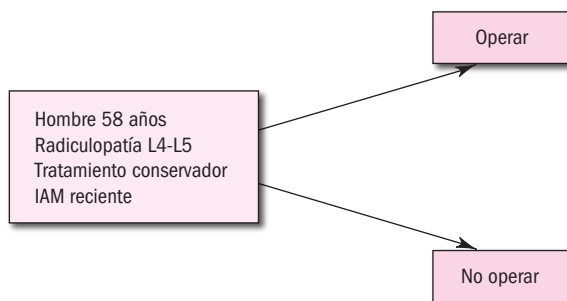


**Figura 14-7.** Identifique el problema y las alternativas de solución.

el nodo de decisión con la utilidad más alta esperada representa la mejor alternativa para el mayor número de enfermos, pero nunca significa que todos ellos tendrán un resultado satisfactorio.<sup>2,27</sup>

El paciente del caso clínico tiene una indicación operatoria muy precisa, pero puede presentar riesgo quirúrgico muy elevado (figura 14-7).

Como segundo paso, se consulta la literatura y se encuentra lo siguiente: después de la laminectomía, el enfermo puede curarse completamente; es posible que la mejoría sólo sea parcial o quizá el sujeto no mejore en absoluto. En estos dos últimos casos puede ocurrir que el síntoma sea dolor o secuela neurológica, o ambos, ya sea por falta del tratamiento o como consecuencia de complicaciones asociadas con la técnica quirúrgica; además, también es posible que ocurra una complicación grave por la intervención quirúrgica porque recientemente sufrió un infarto al miocardio (de hecho, el individuo puede morir).

El enfermo que no se opera puede curarse espontáneamente, mejorar sólo de manera parcial o no sanar en absoluto.

En esta etapa, ya se está en condiciones de identificar las consecuencias de las alternativas de acción (figura 14-8) que se plantearon previamente.

Como tercer paso reconsidera que el riesgo anestésico de la laminectomía en un individuo joven es menor de 0.2%, pero el enfermo con infarto al miocardio reciente se encuentra en la categoría III de Goldman y esto significa riesgo de 13% de sufrir complicaciones cardiovasculares serias durante la anestesia (muerte, infarto, arritmia o embolia pulmonar).<sup>57</sup> La laminectomía, según Deyo, mejora 90% de los casos (15% presenta mejoría parcial y 75% resolución completa, mientras que 10% no obtiene mejoría). La laminectomía es especialmente valiosa para reducir el dolor y un poco menos efectiva para disminuir las secuelas neurológicas (que, por otro lado, no suelen ser muy frecuentes).<sup>58</sup>

También se sabe que 50% de los enfermos se cura espontáneamente sin operación y que 30% no sana en forma espontánea pero presenta algún tipo de mejoría parcial; el resto continúa igual o empeora.<sup>59</sup>

Ahora se puede colocar esta información en el árbol de decisión y es posible asignar la probabilidad con la que ocurrirá cada una de las consecuencias de las alternativas de acción (figura 14-9). ¿Se puede tomar una decisión en este momento?

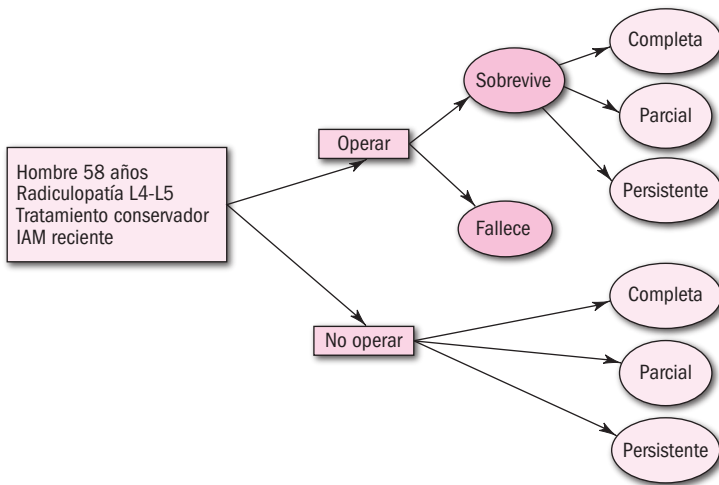


**Figura 14-8.** Identifique las consecuencias de cada alternativa.

Deben anotarse los porcentajes arriba de cada nodo de probabilidad; ha de recordarse que en cada uno la probabilidad de sucesos subsecuentes debe sumar 1.0 (complicaciones 0.13 + sobrevivientes 0.87 = 1.0 y así sucesivamente).

Como cuarto paso se estima la probabilidad de que ocurra cada consecuencia. Para calcular la probabilidad de mejoría parcial después de laminectomía, se multiplica 0.15 (mejoría parcial)  $\times$  0.87 (probabilidad de soportar la intervención quirúrgica sin com-

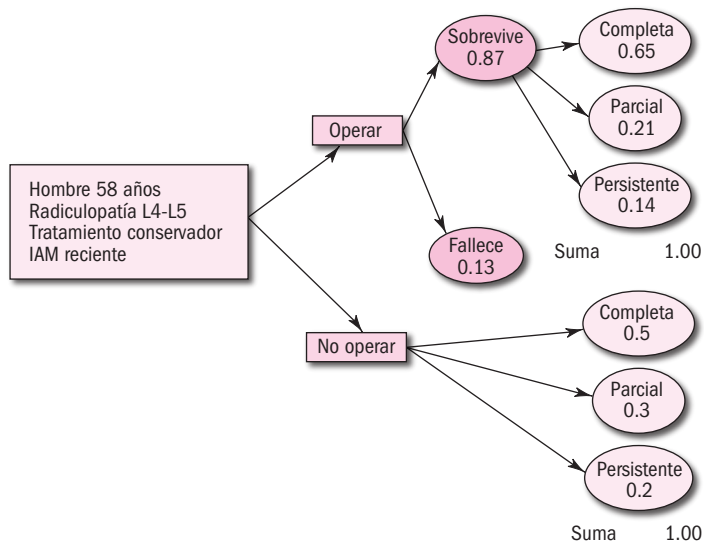


**Figura 14-9.** Identifique las probabilidades con que ocurre cada consecuencia.

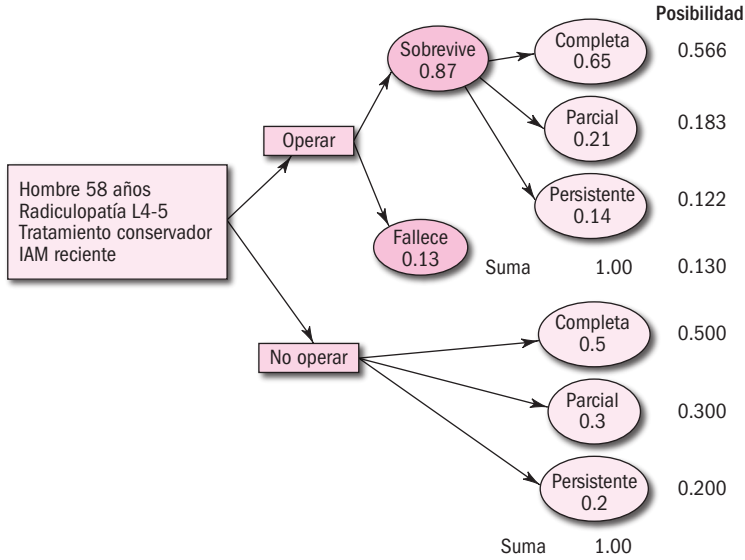


**Figura 14-10.** Calcule con qué probabilidad ocurriría cada consecuencia, multiplicando por las probabilidades anteriores.

plicaciones); el resultado es 0.13, lo que significa que 13% de los casos obtiene mejoría parcial sin complicaciones mayores.

Se hace lo mismo para cada rama; recuerde que el resultado de cada nodo de decisión (operar, no operar) debe sumar 1.0.

Ahora calcule con qué probabilidad ocurrirá cada resultado. Note que en el caso de la opción de operar, existen dos niveles: la probabilidad de sobrevivir (0.87) y la de cada resultado. Para calcular la probabilidad de sobrevivir y de cada uno de esos resultados, debe multiplicar “hacia atrás”; es decir, cada resultado se multiplica por la probabilidad de sobrevivir (figura 14-10).

Como sexto paso se considera que el resultado ideal es curar y el peor posible lo constituye el desarrollo de complicaciones (infarto, embolia pulmonar, arritmia o muerte). Al primer caso puede asignársele un valor de 1.0, y al segundo, un valor de 0.0; ¿qué valores se asignaron a los dos resultados intermedios?

En este ejemplo se usan estimaciones de utilidad que elaboró un grupo de estudiantes de Medicina; note que para ellos es casi tan aceptable la mejoría parcial (a decir de los estudiantes, porque reduce el dolor) que la curación completa, en tanto que la posibilidad de no mejoría tiene un impacto muy importante (50% de reducción en la preferencia de dicho grupo en relación con el resultado ideal). Anote el resultado a la derecha del cálculo de probabilidad de cada rama (figura 14-11).

Como séptimo paso se multiplica cada uno de los valores de probabilidad por el de utilidad correspondiente (en sentido horizontal) y se suman todos éstos en sentido vertical dentro de cada nodo de decisión (figura 14-12).

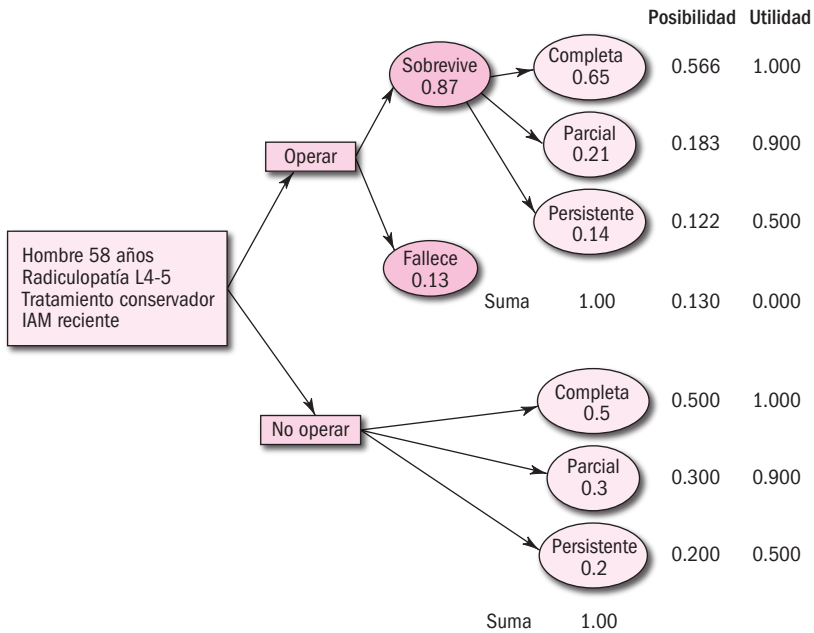


**Figura 14-11.** Incorpore la utilidad en cada consecuencia.

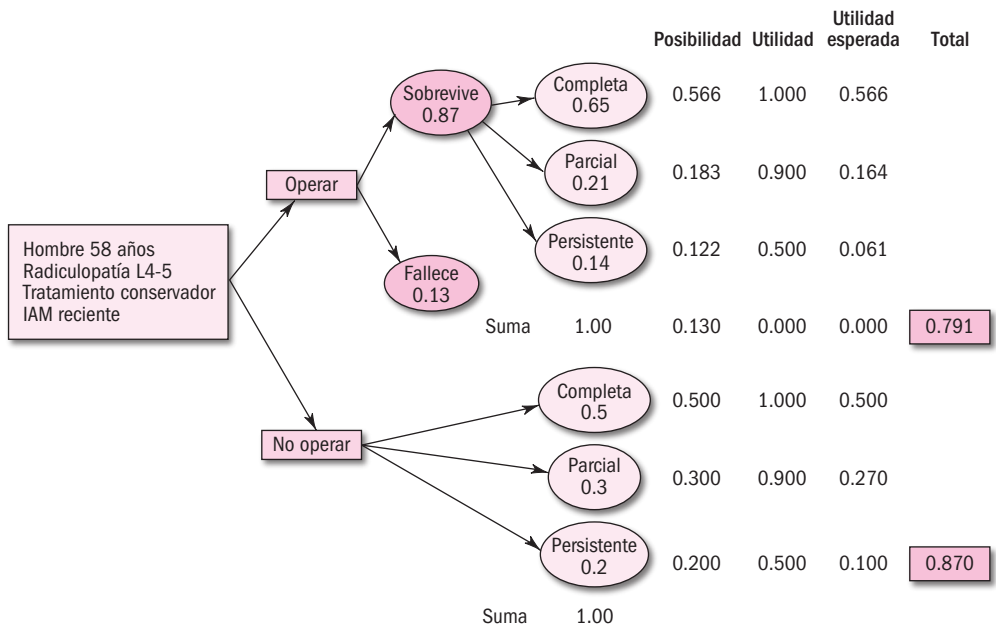


**Figura 14-12.** Calcule la utilidad esperada, multiplicando la probabilidad por la utilidad y sume las correspondientes a cada nodo de decisión.

Como octavo paso, se comparan los resultados de las operaciones anteriores. ¿Cuál fue mayor? En el caso del ejemplo, se obtiene un valor de 0.812 para la intervención quirúrgica y 0.87 para la opción de no operar. En promedio se lograrán mejores resultados si no se opera el enfermo. ¿Fue esta la opinión del lector desde el principio? ¿La diferencia es del tamaño que se esperaba?

Si el lector tiene alguna experiencia y conoce la enfermedad y el riesgo quirúrgico de un caso similar, sin duda su respuesta a la primera pregunta fue afirmativa. Es posible esperar algunas semanas; la probabilidad de que el individuo mejore espontáneamente es muy alta y no se lo somete al riesgo quirúrgico. Si para entonces el sujeto no mejoró lo suficiente, entonces puede proponérsele la operación con menor riesgo. Así, por ejemplo, si el riesgo quirúrgico es de 0.7% (categoría 1 de Goldman), la utilidad esperada de la opción quirúrgica asciende a 0.93, mientras que la opción no quirúrgica continúa siendo 0.87.

¿Qué ocurrirá si la utilidad de la mejoría parcial se reduce a 0.75 y la de no mejoría a 0.25? Repita el análisis. ¿Cuál es su resultado? Intente responder estas preguntas (figura 14-13).

### Ejemplo de proceso de Markov

Considere el caso de un hombre de 55 años de edad con diabetes tipo 2 de reciente diagnóstico que inicia tratamiento con metformina, 850 mg cada 12 h, sin complicaciones crónicas. Usted desea conocer cuál es la probabilidad de desarrollar nefropatía diabética a lo largo de su enfermedad.

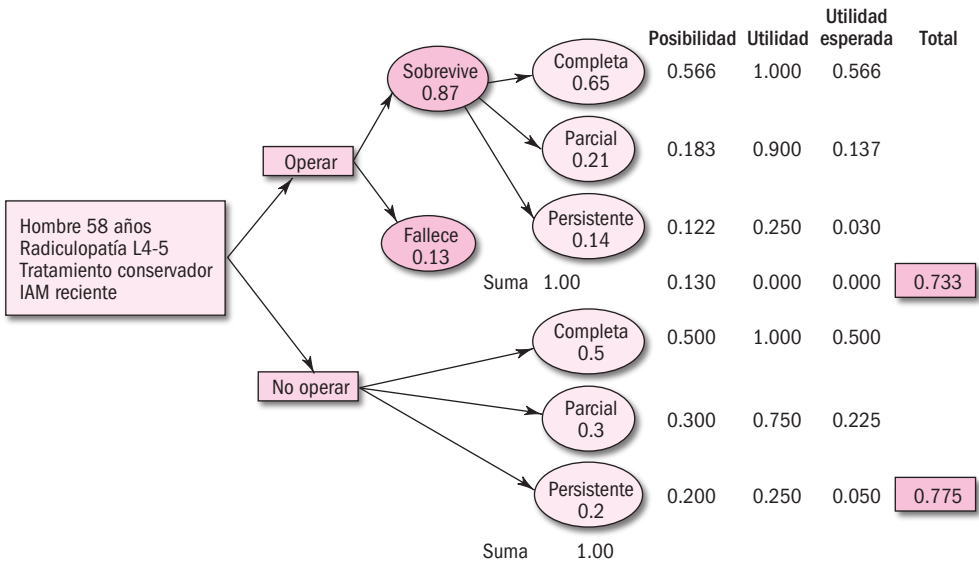


**Figura 14-13.** Análisis de decisión: cambie las condiciones del análisis; en este caso, se variaron los estimados de utilidad. Recalcule la utilidad esperada. Si en todos los casos uno de los nodos de decisión arroja valores superiores al otro, tiene usted una estimación robusta y puede tomar su decisión. Si no, tendrá que hacer un análisis de gatillo, como el que se muestra en la figura 14-5.



Así construirá el modelo de nefropatía diabética, mediante definir inicialmente los estados de salud como cada una de las etapas de esta complicación: normoalbuminuria, microalbuminuria, IRC/diálisis y muerte. Obtiene las probabilidades de transición de lo reportado en la literatura; en este caso, utiliza lo reportado por el estudio UKPDS,<sup>60</sup> después de 20 años de seguimiento. Ahí se informa que un paciente que inicia la diabetes en normoalbuminuria tiene una probabilidad anual de desarrollar microalbuminuria de 2.0%; de presentar macroalbuminuria asciende a 2.8%; si progresa a IRC de 2.3%, y de morir por nefropatía diabética de 1.4%. Además, por tratarse de una enfermedad crónica se establece que la duración de cada ciclo será de un año (figura 14-14).

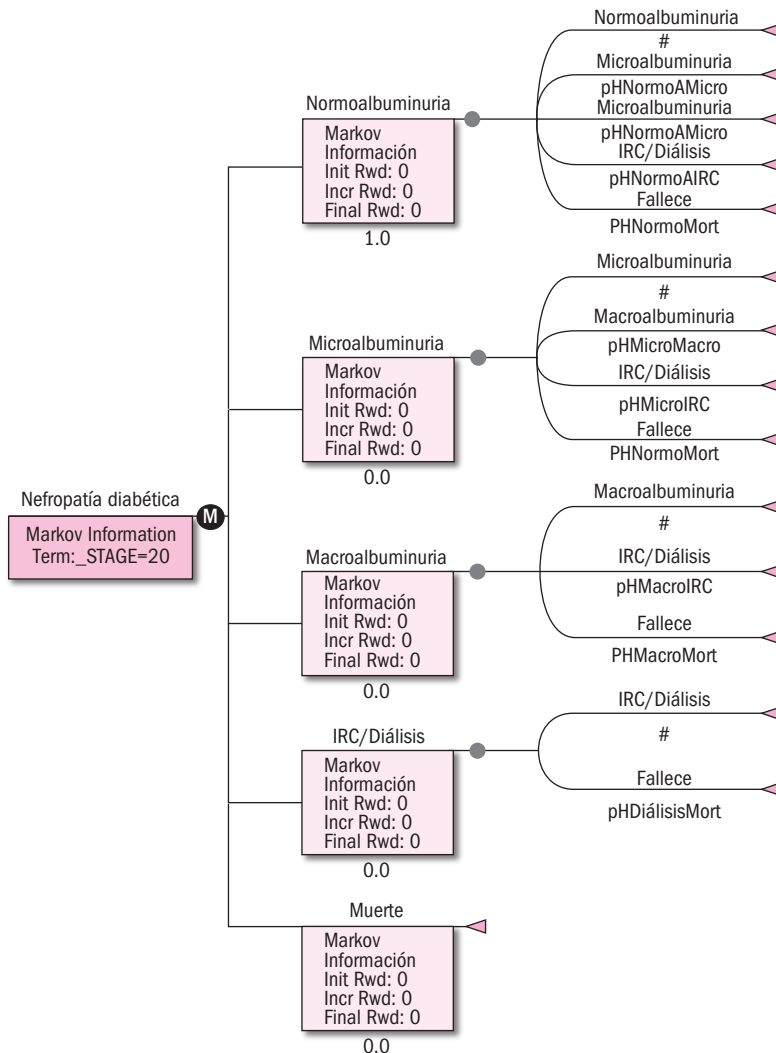


**Figura 14-14.** Modelo del análisis de Markov para nefropatía diabética.

Note que en cada paso el paciente puede “brincar” categorías, es decir, hay enfermos que se mantienen aparentemente con buena función renal y en el año subsecuente presentan macroalbuminuria o insuficiencia renal crónica. Otros fallecen sin deterioro de la función renal. Imagine cómo habría que hacer el árbol de decisión si en cada nodo se puede “brincar” a todos los demás. El árbol de decisión sería demasiado complejo e imposible de analizar.

En un caso como el antes mencionado, resulta más conveniente proponer un análisis de Markov. En este modelo hay cinco estados (estados de Markov) y una serie posible de transiciones. Advierta cómo desde normoalbuminuria se puede pasar a cualesquiera de esos estados; pero la probabilidad de transición más alta es a microalbuminuria; los que siguen a microalbuminuria ocurren con probabilidades muy bajas. Desde microalbuminuria pueden ocurrir los demás estados, menos normoalbuminuria, con probabilidades más altas para macroalbuminuria y más bajas para las demás. El último estado es la muerte.

Este análisis se repite por 20 veces (ciclos anuales) y la figura 14-15 muestra las probabilidades acumuladas de cada evento en cada uno de los 20 años. Si usted multiplica todos estos valores por 1 000, sabría cuántos pacientes de una “cohorte hipotética” estarían en cada una de las condiciones al cabo de 20 años.

Aquí podría agregar el cálculo de utilidades: multiplique por 0.75 cada año de supervivencia con normoalbuminuria, microalbuminuria y macroalbuminuria por 0.4 por cada

| Año                                                                                                                                              | Normal | Micro | Macro | IRC/diálisis | Muerte |
|--------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|-------|--------------|--------|
| 0                                                                                                                                                | 1.000  | 0.000 | 0.000 | 0.000        | 0.000  |
| 1                                                                                                                                                | 0.964  | 0.020 | 0.001 | 0.001        | 0.014  |
| 2                                                                                                                                                | 0.929  | 0.038 | 0.002 | 0.002        | 0.028  |
| 3                                                                                                                                                | 0.896  | 0.054 | 0.004 | 0.003        | 0.043  |
| 4                                                                                                                                                | 0.864  | 0.069 | 0.006 | 0.003        | 0.058  |
| 5                                                                                                                                                | 0.833  | 0.082 | 0.009 | 0.004        | 0.073  |
| 6                                                                                                                                                | 0.803  | 0.094 | 0.011 | 0.004        | 0.088  |
| 7                                                                                                                                                | 0.774  | 0.104 | 0.014 | 0.005        | 0.104  |
| 8                                                                                                                                                | 0.746  | 0.113 | 0.017 | 0.005        | 0.119  |
| 9                                                                                                                                                | 0.719  | 0.121 | 0.019 | 0.006        | 0.135  |
| 10                                                                                                                                               | 0.693  | 0.128 | 0.022 | 0.006        | 0.150  |
| 11                                                                                                                                               | 0.668  | 0.134 | 0.025 | 0.007        | 0.166  |
| 12                                                                                                                                               | 0.644  | 0.139 | 0.028 | 0.007        | 0.182  |
| 13                                                                                                                                               | 0.621  | 0.144 | 0.030 | 0.007        | 0.198  |
| 14                                                                                                                                               | 0.599  | 0.147 | 0.033 | 0.008        | 0.214  |
| 15                                                                                                                                               | 0.577  | 0.150 | 0.035 | 0.008        | 0.229  |
| 16                                                                                                                                               | 0.556  | 0.153 | 0.038 | 0.008        | 0.245  |
| 17                                                                                                                                               | 0.536  | 0.155 | 0.040 | 0.009        | 0.261  |
| 18                                                                                                                                               | 0.517  | 0.156 | 0.042 | 0.009        | 0.276  |
| 19                                                                                                                                               | 0.498  | 0.157 | 0.044 | 0.009        | 0.292  |
| 20                                                                                                                                               | 0.480  | 0.157 | 0.046 | 0.009        | 0.307  |
| Abreviaturas: Normal = normoalbuminuria, Micro = microalbuminuria, Macro = macroalbuminuria, IRC/diálisis = insuficiencia renal crónica/diálisis |        |       |       |              |        |

**Figura 14-15.** Estimación de probabilidades al cabo de 20 años (20 ciclos) para el modelo señalado en la figura 14-14.

año en IRC/diálisis y por 0 para muerte. Sume cada renglón en una nueva columna a la derecha y nuevamente sume esa columna en sentido vertical. ¿Cuántos años de vida ajustados por calidad obtiene? (consulte el resultado en la figura 14-16).

Todas estas operaciones se pueden llevar a cabo en una hoja de cálculo en su computadora personal. Usted podría comparar resultados si cambia las probabilidades de transición o los valores de la “utilidad” y así habría hecho un análisis de sensibilidad igual al que se hizo para el árbol de decisión.

## Conclusión

Las técnicas de análisis de decisión son herramientas útiles para comparar alternativas de acción, estimar cuantitativamente las posibles consecuencias de una decisión e incorporar las preferencias del paciente en el proceso de decisión y calcular el efecto que la incertidumbre pueda tener en el resultado final. Éste depende de que se construya un modelo apropiado y se obtengan estimados certeros de las probabilidades con que ocurre

| Año   | Normal | Micro | Macro | IRC/diálisis | Muerte | AVAC   |
|-------|--------|-------|-------|--------------|--------|--------|
| 0     | 1.000  | 0.000 | 0.000 | 0.000        | 0.000  | 0.750  |
| 1     | 0.964  | 0.020 | 0.001 | 0.001        | 0.014  | 0.739  |
| 2     | 0.929  | 0.038 | 0.002 | 0.002        | 0.028  | 0.728  |
| 3     | 0.896  | 0.054 | 0.004 | 0.003        | 0.043  | 0.717  |
| 4     | 0.864  | 0.069 | 0.006 | 0.003        | 0.058  | 0.705  |
| 5     | 0.833  | 0.082 | 0.009 | 0.004        | 0.073  | 0.695  |
| 6     | 0.803  | 0.094 | 0.011 | 0.004        | 0.088  | 0.683  |
| 7     | 0.774  | 0.104 | 0.014 | 0.005        | 0.104  | 0.671  |
| 8     | 0.746  | 0.113 | 0.017 | 0.005        | 0.119  | 0.659  |
| 9     | 0.719  | 0.121 | 0.019 | 0.006        | 0.135  | 0.647  |
| 10    | 0.693  | 0.128 | 0.022 | 0.006        | 0.150  | 0.635  |
| 11    | 0.668  | 0.134 | 0.025 | 0.007        | 0.166  | 0.623  |
| 12    | 0.644  | 0.139 | 0.028 | 0.007        | 0.132  | 0.611  |
| 13    | 0.621  | 0.144 | 0.030 | 0.007        | 0.198  | 0.599  |
| 14    | 0.599  | 0.147 | 0.033 | 0.008        | 0.214  | 0.587  |
| 15    | 0.577  | 0.150 | 0.035 | 0.008        | 0.229  | 0.575  |
| 16    | 0.556  | 0.153 | 0.038 | 0.008        | 0.245  | 0.563  |
| 17    | 0.536  | 0.155 | 0.040 | 0.009        | 0.261  | 0.552  |
| 18    | 0.517  | 0.156 | 0.042 | 0.009        | 0.276  | 0.540  |
| 19    | 0.498  | 0.157 | 0.044 | 0.009        | 0.292  | 0.528  |
| 20    | 0.480  | 0.157 | 0.046 | 0.009        | 0.307  | 0.516  |
| Total |        |       |       |              |        | 13.322 |

Abreviaturas: normal = normoalbuminuria, Micro = microalbuminuria, Macro = macroalbuminuria, IRC/diálisis = insuficiencia renal crónica/diálisis AVAC = Años de vida ajustados por calidad (“QALY”). Éstos se obtienen al multiplicar los casos normal, micro y macro por 0.75; IRC por 0.40 y muerte por 0 (cero). El resultado final de cada año se suma en la casilla inferior. Esto significa que en cada 1000 pacientes, al cabo de 20 años obtendría 13.322 años de vida ajustados por calidad.

**Figura 14-16.** Años de vida ajustados por calidad, incluyendo valores de utilidad en la tabla de probabilidades que se muestra en la figura 14-15.

cada evento y de las preferencias de los pacientes. Son especialmente útiles cuando el modelo es complejo y no es factible llegar a una decisión de manera intuitiva.

El clínico debe tener presente en todo momento que la decisión final y el curso de acción que tome en un caso determinado es de su propia responsabilidad y que el modelo de análisis que elija sólo es una herramienta.

## Referencias

1. McClamroch K, Kaufman J, Behets F. A formal decision analysis identifies an optimal treatment strategy in a resource-poor setting. *J Clin Epidemiol* 61:776-787. 2008.
2. Sox H, Blatt M *et al.* Medical decision making. Butterwords, Stoneham, EUA, 1-64. 1988.
3. Lifshitz A, Quiñónez A. La simplificación postecnológica en la medicina clínica. *Ciencia y Desarrollo* 15:19-23. 1989.
4. Feinstein A. An additional basic science for clinical medicine: III The challenges of comparison and measurement. *Ann Intern Med* 99:705-712. 1983.
5. Sacket D, Haynes B, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Little Brown, Boston, EUA, 3-158. 1985.
6. Kassir J. Diagnostic reasoning. En: Sox H. *Common diagnostic tests: use and interpretation*, 2a. ed. American College of Physicians, Filadelfia, EUA, 1-15. 1990.
7. McNeil B, Keeler E, Adelstein J. Primer on certain elements of medical decision making. *N Engl J Med* 293:211-215. 1975.
8. Nettleman M. Receiver operator characteristic (ROC) curves. *Infect Cont Hosp Epidemiol* 9: 374-377. 1988.
9. Sox H. *Probability theory and the interpretation of diagnostic tests*, 2a. ed. American College of Physicians, Filadelfia, EUA, 16-33. 1990.
10. Feinstein A. Invidious comparisons and unmet clinical challenges. *JAMA* 92:117-120. 1992.
11. Theodossi A, Spiegelhalter D, McFarlane I, Williams R. Doctors' attitudes to risk in difficult clinical decisions: application of decision analysis in hepatobiliary disease. *Br Med J* 289:213-216. 1984.
12. Tversky A, Kahneman D. Judgement under uncertainty: heuristics and biases. *Science* 185:1124-1131. 1974.
13. Tversky A, Kahneman D. The framing of decisions and the rationality of choice. *Science* 211: 453-458. 1981.
14. Schmidt H, Norman G, Boshuizen H. A cognitive perspective on medical expertise: theory and implications. *Acad Med* 65:611-621. 1990.
15. Feinstein A. Clinical judgement. Psychological research and medical practice. *Science* 194: 696-700. 1976.
16. Calva J, Ponce de León S *et al.* Cómo leer revistas médicas. *Rev Invest Clin* 40:65-106. 1988.
17. O'Neill P, Devlin N. An analysis of NICE's 'restricted' (or 'optimized') decisions. *Pharmacoeconomics* 28:987-993. 2010.
18. Kassirer J, Kopelman R. Memory: how we overcome its imitations. *Hosp Pract* 23:25-29. 1988.
19. Pauker S, Kassirer J. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med* 293: 229-234. 1975.
20. McNeil B, Pauker S. The patient's role in assessing the value of diagnostic tests. *Radiology* 132:605-610. 1979.

21. Pauker S, Kassirer J. Decision analysis. *N Engl J Med* 316:250-258. 1987.
22. Henschke U, Flehinger B. Decision theory in cancer therapy. *Cancer* 20:1819. 1967.
23. Ransohoff D, Gracie W *et al.* Prophylactic cholecistectomy or expectant management for silent gallstones: a decision analysis to assess survival. *Ann Intern Med* 99:1999-1204. 1983.
24. Kassirer J. Is renal biopsy necessary for optimal management of idiopathic nephrotic syndrome? *Kidney Int* 24:56-75. 1983.
25. Taylor W, Aronson M, Delbanco T. Should young adults with a positive tuberculin test take isoniazid? *Ann Intern Med* 94:808-813. 1981.
26. McNeil B, Weichselbaum R, Pauker S. Speech and survival: trade-off between quality and quantity of life in laryngeal cancer. *N Engl J Med* 305:982-987. 1984.
27. Weinstein M, Fineberg H. *Clinical decision analysis*. WB Saunders, EUA, 1980.
28. Grobman W. Decision analysis in obstetrics and gynecology. *Obstetrical and Gynecological Survey* 61:602-607. 2006.
29. Weinstein M, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce B. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health* 6:9-17. 2003.
30. Kassirer J, Moskowitz A, Lau J. Decision analysis: a progress report. *Ann Intern Med* 106:275-282. 1987.
31. Epstein P. Cassandra and the clinician: are clinical prediction rules changing the practice of medicine? *Ann Intern Med* 113:646-647. 1990.
32. Lifshitz A. Sobre las asociaciones de las enfermedades. *Gac Med Mex* 124:119-126. 1988.
33. Reilly B, Evans A. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 144:201-209. 2006.
34. Moskowitz A, Kuipers B, Krassier J. Dealing with uncertainty, risks, and trade-offs in clinical decisions. A cognitive science approach. *Ann Intern Med* 108:435-439. 1988.
35. Berwick D, Fineberg H, Weinstein M. When doctors meet numbers. *JAMA* 271:991-998. 1981.
36. Pauker S, Kassirer J. The threshold approach to clinical decision making. *N Engl J Med* 302:1109-1117. 1980.
37. Rubio-Terrés C. Introducción a la utilización de los modelos de Markov en el análisis farmacoeconómico. *Farm Hosp* 24:241-247. 2000.
38. Beck J, Pauker S. The Markov process in medical prognosis. *Med Des Marking* 3:419-458. 1983.
39. Kymes S. An introduction to decision analysis in the economic evaluation of the prevention and treatment of vision-related diseases. *Ophthalm Epidemiol* 15:76-83. 2008.
40. Hilgsmann M, Ethgen O, Bruyère O, Richy F, Gathion H, PhD, Reginster J. Development and validation of a Markov Microsimulation model for the economic evaluation of treatments in osteoporosis. *Value in Health* 12:687-696. 2009.
41. Geisler B, Siebert U, Gazelle G, Cohen D, Göhler A. Deterministic sensitivity analysis for first-order Monte Carlo simulations: a technical note. *Value in Health* 12:96-97. 2009.
42. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. *Value in Health* 13:557-564. 2010.
43. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health* 12:739-749. 2009.
44. Guo J, Pandey S, Doyle J, Bian B, Lis Y, Raisch D. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy —Report of the ISPOR Risk-Benefit Management Working Group. *Value in Health* 13:657-666. 2010.

45. Boshuizen H, van Baal P. Probabilistic sensitivity analysis: be a Bayesian. *Value in Health* 12: 1210-1214. 2009.
46. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? *Value in Health* 9:244-252. 2006.
47. Protheroe J, Fahey T, Montgomery A, Peters T. The impact of patients' preferences on the treatment of atrial fibrillation: observational study of patient based decision analysis. *Br Med J* 320:1380-1384. 2000.
48. Richardson G, Bojke C, Kennedy A *et al.* What outcomes are important to patients with long term conditions? A discrete choice experiment. *Value in Health* 12:331-339. 2009.
49. Say R, Thomson R. The importance of patient preferences in treatment decisions—challenges for doctors. *Br Med J* 327:542-545. 2003.
50. Torrance W. Utility approach to measuring health related quality of life. *J Chronic Dis* 40: 593-600. 1987.
51. Forberg D, Kane R. Methodology for measuring health state preferences-II: scaling methods. *J Clin Epidemiol* 42:459-471. 1989.
52. Dawes R, Faust D, Meehl P. Clinical *versus* actuarial judgement. *Science* 243:1668-1674. 1989.
53. Feinstein A. Scientific standards in epidemiology studies of the menace of daily life. *Science* 242:1257-1263. 1988.
54. Forrow L, Taylor W, Arnold R. Absolutely relative: how research results are summarized can affect treatment decisions. *JAMA* 121-124. 1992.
55. Detsky A, Redelmeier D, Abrams H. What's wrong with decision analysis? Can the left brain influence the right? *J Chronic Dis* 40:831-836. 1987.
56. Critchfield G, Willard K. Probabilistic analysis of decision trees using Monte Carlo simulation. *Med Des Marking* 6:85-92. 1986.
57. Goldman L, Caldera D, Nussbaum S *et al.* Multifactorial index of cardiac risk in non-cardiac surgical procedures. *N Engl J Med* 297:845-850. 1977.
58. Deyo L, Loeser J, Bigos S. Herniated lumbar intervertebral disk. *Ann Intern Med* 112:598-603. 1990.
59. Weber H. Lumbar disc herniation: a controlled prospective study with ten years of observation. *Spine* 8:131-139. 1983.
60. Adler AI, Stevens RJ, Manley SE, Bilous RW, Cull CA, Holman RR. Development and progression of nephropathy in type 2 diabetes: The United Kingdom Prospective Diabetes Study UKPDS 64. *Kidn Int* 63:225-232. 2003.

---

## Bibliografía

- Kocher M, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Joint Surg* 86A:607-620. 2004.
- Pauker S, Mc Neil B. Impact of patient preferences on the selection of therapy. *J Chronic Disc* 34:77-86. 1981.
- Rodríguez-Pimentel L, Silva-Romo R, Wachter N. Estudios económicos y análisis de decisiones en la autogestión de recursos. *Rev Med Inst Mex Seguro Soc* 45:297-304. 2007.
- Wasson J, Harold H *et al.* Clinical prediction rules: applications and methodological standards. *N Eng J Med* 313:793-799. 1985.



# Capítulo 15

## Calidad de vida: aproximaciones a su medición

*Miguel Ángel Villasís Kever  
Jesús Arias Gómez*

### Introducción

La salud es un elemento importante para los individuos, como lo demuestra la gran cantidad de recursos que se invierten tanto en su cuidado como en su recuperación. Disfrutar de un buen estado de salud es un derecho incuestionable de los individuos, y para garantizarlo se emprenden acciones desde diferentes ámbitos mediante la legislación, educación, modificación del medio y otras acciones tendientes a mejorar la salud de los individuos y las comunidades.

Desde un punto de vista epidemiológico, la esperanza de vida es uno de los indicadores tradicionales que pueden ayudar a conocer las modificaciones en cuanto a las condiciones de salud de las poblaciones; los países con mayor esperanza de vida tienen mejores condiciones de salud. Lo mismo ocurre con otros indicadores como el índice de natalidad y mortalidad, o bien, la prevalencia o incidencia de las enfermedades. Sin embargo, esto no refleja la opinión o la percepción de los individuos en cuanto a su estado de salud-enfermedad.

La información que proporciona medir esta percepción permite identificar distintos estados de morbilidad y bienestar y evaluar el impacto de las intervenciones terapéuticas y sanitarias, así como determinar la forma en que repercute el estado de salud-enfermedad en el ámbito físico, psicológico y social de los individuos. La salud autopercibida o calidad de vida se asocia de manera independiente al diagnóstico clínico y ayuda a predecir la utilización de los servicios de atención médica.

Para el área médica, la medición de la calidad de vida en relación con la salud es un tema creciente, ya que se ha determinado que es necesario medir el impacto de las acciones dirigidas al cuidado de los pacientes desde una perspectiva más individual, pero su evaluación no siempre es sencilla o concreta. Tradicionalmente se han empleado medidas indirectas y negativas, es decir, en lugar de medir la “salud”, se mide la “falta de salud”,



como la tasa de mortalidad, el cual es un ejemplo de indicador indirecto y negativo, ya que mide la falta de salud a través de la muerte.

Otros indicadores, como las tasas de morbilidad y esperanza de vida, miden el resultado de los cuidados para la salud desde la perspectiva de los profesionales encargados de la misma, pero adolecen de la opinión de los pacientes (quienes son objeto de los cuidados), por lo que se considera que la evaluación de las intervenciones es incompleta, ya que son los sujetos intervenidos quienes deben opinar sobre la forma en que sus vidas o su salud son afectadas.

Para contextualizar estos conceptos podemos pensar en los pacientes con cáncer; los tratamientos más recientes han ayudado a incrementar el número de sobrevivientes; sin embargo, estos tratamientos ocasionan más efectos adversos, que pueden ser graves y hacer que los pacientes se sientan muy mal. Otro ejemplo son los pacientes con alguna discapacidad, congénita o adquirida, en quienes se podría asumir que ante la falta de funcionalidad se sientan mal, pero se ha demostrado que esto no necesariamente es cierto cuando se les pregunta directamente. Por lo anterior, se considera que para tener una evaluación integral del estado de salud de los individuos o los posibles cambios que ocurren por las intervenciones, es necesario disponer tanto de datos duros (incremento en los meses o años de la vida) como de datos cualitativos, es decir, la percepción de los individuos de su calidad de vida.

## Caso clínico

---

Se presenta una paciente femenina de 49 años, casada, dedicada al hogar, y quien terminó la preparatoria. En la actualidad tiene dos hijos y hasta antes del inicio de su enfermedad realizaba aproximadamente una hora diaria de ejercicio. Hace alrededor de un año inició con rigidez, edema y dolor matutino de articulaciones de ambas manos. La sintomatología ha sido progresiva, observándose discreta deformación de estas articulaciones. Desde hace seis meses se agregó afección de rodillas y codos, lo cual limita más sus actividades diarias.

Acudió con un médico internista, quien diagnosticó artritis reumatoide, por lo que inició con diclofenaco y, ante la persistencia de los síntomas, agregó metotrexato. Sin embargo, dos meses después persistió con la inflamación y ha progresado la limitación en sus actividades físicas.

La paciente, además, refiere presentar cansancio fácil, tristeza y cambios de talante; esto último lo relaciona con su incapacidad física. Ante la progresión, se decide probar la terapia con rituximab, que es un medicamento que recientemente ha probado su eficacia en ensayos clínicos controlados en este tipo de pacientes.

El médico internista, antes de iniciar el nuevo tratamiento, considera importante evaluar con mayor precisión las modificaciones que puedan resultar del cambio de terapéutica. Dentro de las evaluaciones, decide incluir la percepción por parte de la paciente. Sin embargo, le surgen las siguientes interrogantes: ¿cómo evaluar la percepción de la paciente de manera objetiva?, ¿será necesario utilizar algún instrumento o escala que mida la calidad de vida?, ¿existirá algún cuestionario en español que sea sencillo de llenar por la paciente?, ¿cómo interpretar los resultados?

---



## Enfoque histórico de la calidad de vida

El estudio de la salud, del bienestar, la felicidad y de la calidad de vida es tan antiguo como los seres humanos. Se ha identificado que en civilizaciones antiguas como la egipcia, hebrea, griega y romana, existía una preocupación por la salud de las personas, considerada como un espacio público y colectivo. Esta tradición continuó en el Imperio Bizantino y se extendió hacia el mundo árabe; más tarde, en la Edad Media se introdujo en Europa, pero con retrocesos debido al abandono de los componentes privado y público, pero con un cambio de conductas en relación con la higiene. Obviamente, el alcance de estas medidas fue limitado y local; sin embargo, se establecieron una serie de normas acerca de la salud pública con respecto a la contaminación del agua, el estado de los alimentos en los mercados, alcantarillado, pavimentación y limpieza de calles.

Junto con la salubridad, comienza a hablarse de condiciones de trabajo, dieta y sistema alimentario, de tasas de mortalidad y esperanza de vida, de sistemas de escolarización, de las condiciones de la vivienda y las ciudades y de las formas de vida de las distintas clases sociales. De esta manera, a finales del siglo XVIII y principios del XIX aparecieron las legislaciones acerca de la regulación del trabajo y el descanso, de la enfermedad y los accidentes, de la maternidad, de la vejez y de la muerte, por lo que nace el área de bienestar social. Con el reconocimiento de los derechos de los ciudadanos en los estados, la calidad de vida adquiere una expansión social y política.

Desde el punto de vista ideológico, la raíz de la investigación de la calidad de vida nace en el siglo XVIII con el pensamiento generado por la época de la Ilustración, donde el propósito de la vida humana es la vida por sí misma, adoptando a la autorrealización y la felicidad como los valores centrales de esa época. En el siglo XIX la sociedad se preocupa por proporcionar una buena vida a sus ciudadanos, y es a través de la creación del **utilitarismo** como se busca una mejor sociedad, en la cual se proporcione “más felicidad al mayor número de ciudadanos”. En el siglo XX, con el enfoque previo, se intenta planear una reforma social del estado de bienestar social. Con el impulso de las ciencias sociales en las décadas de 1950-1959 y 1960-1969 cobra auge el interés por el bienestar social y su medición; a esta época se le conoció como el **movimiento de los indicadores sociales**, los cuales se caracterizaron por ser medidas objetivas y normativas de una población o grupo determinado. Es en esta época cuando se acuñó el término **calidad de vida**, que algunos autores atribuyen al economista estadounidense John Kenneth Galbraith a finales del decenio de 1950-1959 y otros al presidente estadounidense Lyndon B. Johnson en 1964. Un poco más adelante se comenzó a considerar que las condiciones sociales producen cambios en la percepción de la calidad de vida. Al integrar esta otra característica, entonces el concepto de **calidad de vida** amplió su composición, integrando aspectos de bienestar personal y social.

A finales de la década de 1960-1969 y durante la de 1970-1979 el concepto de **calidad de vida** incorporó medidas subjetivas, es decir, evaluaciones que incluían aspectos como los sentimientos personales, como la satisfacción, la sensación de bienestar y la felicidad. Desde entonces, el concepto de **calidad de vida** ha evolucionado de manera

significativa, ya que implicó entender que para su evaluación no era suficiente disponer de elementos objetivos, sino que se debería incluir una perspectiva subjetiva. Este nuevo enfoque le dio una entidad propia, diferenciando la calidad de vida de otros indicadores sociales, poniendo énfasis en lo que los sujetos perciben sobre su vida y sobre las condiciones objetivas de su existencia. Los avances en los años más recientes sobre calidad de vida han sido sobre la mejoría en la comprensión del concepto, su importancia en el contexto de la salud de los individuos sanos y enfermos, así como en el desarrollo de instrumentos para medirla.

## Concepto de “calidad de vida”

La calidad de vida debe reconocerse como un concepto multidimensional que incluye el estilo de vida, la vivienda, satisfacción en la escuela y el empleo, así como la situación económica. Por esta razón, la calidad de vida se conceptualiza de acuerdo con un sistema de valores, estándares o perspectivas que varían de persona a persona, de grupo a grupo, lugar a lugar, así como en el momento de su evaluación. Entonces, la calidad de vida podría considerarse como la sensación de bienestar que puede ser experimentada por las personas y que representa la suma de sensaciones subjetivas y personales del “sentirse bien”. Sin embargo, no existe consenso respecto al significado del término “calidad de vida”; aunque para cualquier definición, se trata de un término más amplio que el bienestar físico y social, por lo cual el concepto integra diferentes variables, entre las que se incluyen la satisfacción, el bienestar —subjetivo—, la felicidad y el estado de salud (físico, psíquico y social). En general, todas las definiciones de calidad de vida tienen dos elementos comunes: utilizan un enfoque multidimensional y hacen hincapié en la valoración que el individuo hace de su propia vida.

## Calidad de vida relacionada con la salud

El término “calidad de vida” empieza a relacionarse con los cuidados sanitarios a mitad del decenio de 1970-1979, cuando la atención sanitaria —influenciada por los cambios sociales— favorece que se incorporen temas como el alivio de síntomas, la mejoría en las actividades de los individuos, así como el establecimiento de relaciones sociales más óptimas o la autonomía. Con los cambios epidemiológicos, los estándares tradicionales en la salud pública, como la morbilidad o mortalidad, se consideran insuficientes para determinar el bienestar de los individuos, en términos de la salud.

En este contexto adquiere importancia el desarrollo de medidas que incorporen lo que un individuo piensa o siente de sí mismo, es decir, la forma en que un sujeto o un paciente percibe su estado de salud-enfermedad. Con este nuevo paradigma, los cuidados para la salud toman un nuevo giro y se valora el impacto de las enfermedades y de sus tratamientos sobre el bienestar y la satisfacción, así como en las capacidades físicas, psicológicas y sociales de los individuos. Sin embargo, este enfoque es muy amplio y abstracto, lo que hace difícil ponerlo en práctica, de ahí que Patrick y Erickson en 1993 sugirieron

utilizar el término “**calidad de vida relacionada con la salud**” (*health related quality of life*), al cual concibieron como los niveles de bienestar y satisfacción asociados con los acontecimientos que una persona puede padecer como consecuencia de una enfermedad o su tratamiento. Este término evita la medición de aspectos no relacionados directamente con la salud, como la libertad o las condiciones del medio que rodea a los individuos.

Con el empleo de nuevas herramientas diagnósticas y terapéuticas que permiten mejorar o mantener las condiciones clínicas, se asume que existe mejoría tanto en la sobrevida como en el estado de salud de los pacientes con enfermedades crónicas; sin embargo, el incremento en el tiempo de sobrevivir no siempre se correlaciona con el aumento en el bienestar de los individuos, por lo que se requiere que ambas sean evaluadas de manera independiente. Por lo anterior, es importante conocer las características de los pacientes y sus familias, las vivencias y percepciones que tienen para determinar el nivel de calidad de vida relacionada con la salud, a fin de comparar si la condición basal se modificó con la implementación de las nuevas tecnologías. Además, es preciso tomar en cuenta los factores específicos que pudieron influir para lograr o no mejoría en la calidad de vida, ya que es posible que se tenga que apoyar otras áreas, de acuerdo con la realidad de los sujetos, como la presencia de alguna complicación o discapacidad, o bien, los problemas económicos relacionados con la propia enfermedad.

El énfasis creciente sobre la prevención de la enfermedad y la promoción de la salud avala la importancia de la calidad de vida relacionada con la salud como un estándar para medir el desempeño de los sistemas de salud de las naciones, por lo que parece importante apoyar la investigación en esta área. Este estándar debiera ser aplicado más ampliamente, de tal forma que vaya más allá de registrar la incidencia y la prevalencia de las enfermedades, que sólo reflejan la efectividad de la prevención primaria y de los cuidados a la salud. La calidad de vida relacionada con la salud debe ser considerada para evaluar la efectividad de las medidas de prevención secundaria y terciaria, así como para la promoción de la independencia o en la reducción de las ausencias laborales en los pacientes con discapacidad. Por otro lado, la medición de la calidad de vida relacionada con la salud puede identificar nuevas relaciones con algunos factores que auxilien o guíen la toma de decisiones con el propósito de que los individuos obtengan la atención a la salud, acorde con sus necesidades individuales. Tal es el caso de la posibilidad de que los pacientes o sus familias participen directamente en el proceso de decisión de las medidas preventivas o terapéuticas.

En un enfoque más ambicioso, al adoptarlo como estándar de salud nacional, debería apoyar en la evaluación de los sistemas de salud, lo que coadyuvaría a evitar las barreras que existen entre las disciplinas sociales y los servicios médicos.

## Investigación en calidad de vida relacionada con la salud

En las décadas recientes, el estudio de la calidad de vida relacionada con la salud ha logrado captar la atención de muchos investigadores que intentan dar respuesta científica a la necesidad de incluir en las evaluaciones en salud la percepción de los usuarios respecto de su bienestar. Si bien la búsqueda de niveles deseables y sostenibles de mejoramiento de la

calidad de vida es una preocupación que siempre ha estado presente en la historia de la humanidad, es relativamente reciente el interés por conceptualizarla y medirla desde una perspectiva integral, por lo que la investigación en el área de la salud ha cobrado auge, desde los puntos de vista epidemiológico, clínico y económico.

La investigación en la calidad de vida, a diferencia de otros resultados en relación con la salud (como la frecuencia de complicaciones o la mortalidad), es probable que sea más sensible para determinar con mayor precisión todas las consecuencias de las enfermedades y sus intervenciones. Sin embargo, a diferencia de otras variables de desenlace o de resultado “duras”, la mayoría de los componentes que integran la calidad de vida son subjetivos, ya que dependen de la percepción de los individuos, lo que crea un reto para su medición y confiabilidad.

Una manera de ejemplificar esto es tomando el caso de la diabetes, donde una de las variables de resultado más importantes desde el punto de vista clínico son las modificaciones en las cifras de glucemia; sin embargo, es posible que las cifras de glucemia óptimas generen malestar a los pacientes, ya que tienen que cambiar su estilo de vida (hacer ejercicio, llevar dietas estrictas) o tomar ciertos medicamentos que causan efectos adversos, por lo que los pacientes pueden considerar que se sienten mal, o bien, calificar que tienen una mala calidad de vida.

Por lo anterior, en el entendido de que la evaluación de la calidad de vida no puede hacerse de manera directa y que se requiere incorporar una serie de elementos para calificarla de manera integral, entonces se han creado instrumentos en los cuales a través de preguntas o ítems se analiza cada una de las partes del concepto de **calidad de vida**, es decir, en estos reactivos (*ítems*) se explora la percepción de los individuos sobre las variables o características físicas, psicológicas, sociales y funcionales de los pacientes que pueden afectar su calidad de vida. Estos reactivos (después de pasar por un arduo proceso metodológico y estadístico para determinar su validez y confiabilidad, que no es el propósito de este capítulo) integran un cuestionario estandarizado capaz de ser aplicado al grupo específico para el cual fue creado, como para pacientes, mujeres, niños, etcétera. Cada respuesta obtenida por reactivo brinda una puntuación; al final, la suma de las puntuaciones proporcionará la calificación por cada dimensión o de manera global de la calidad de vida de cada individuo.

En la mayoría de los instrumentos de calidad de vida se destacan tres aspectos: 1) la importancia del estado funcional (es decir, la evaluación del estado físico, social y mental de los individuos); 2) la subjetividad de la aproximación para medir el estado de salud, lo cual se realiza mediante la resolución de preguntas, y 3) obtener una puntuación que representa la calificación de paciente para el estado de salud individual o su calidad de vida.

La investigación sobre la calidad de vida no ha sido sencilla debido a que relativamente es de interés reciente y a que es un área donde el instrumental científico para su medición no es único, como lo muestra la gran variedad de instrumentos que se han desarrollado, tanto genéricos como específicos, para evaluarla. Los instrumentos genéricos para medir la calidad de vida se refieren a escalas o cuestionarios que se pueden utilizar para cualquier tipo de poblaciones; en cambio, los específicos están diseñados para su aplicación en

grupos particulares, ya que las modificaciones de la calidad de vida están dadas por las características propias de los pacientes o de las enfermedades. De ahí que existen instrumentos específicos para niños y otros para adultos, así como para enfermedades particulares como el cáncer o los trastornos musculoesqueléticos. Un punto que es importante añadir en este contexto es el que se relaciona con quienes califican la calidad de vida; en un inicio se pensaba que sería suficiente evaluar la calidad de vida desde la perspectiva de los trabajadores de la salud; sin embargo, no siempre un estado funcional dado (como el que se mide con la escala de Karnofsky, creada para pacientes con cáncer) corresponde con el sentir de los individuos.

Los instrumentos para medir calidad de vida se utilizan con tres propósitos principales: 1) como instrumentos discriminantes, a fin de diferenciar los pacientes que tienen mejor y peor, o buena y mala, calidad de vida en un momento determinado; 2) como instrumentos evaluativos, en donde el objetivo es medir los cambios en la calidad de vida en diferentes momentos de la enfermedad, y 3) como instrumentos predictivos, cuando se busca estimar cómo será la calidad de vida, a partir de una medición.

En la actualidad existen numerosos instrumentos para medir la calidad de vida; sin embargo, no existe alguno que haya sido aceptado como un estándar que pueda ser aplicado en cualquier población, por lo que si se pretende utilizar algún instrumento como parte de un estudio de investigación o como parte de la evaluación que se hace de manera cotidiana en la clínica (p. ej., la escala de Karnofsky se utiliza en pacientes con cáncer, aunque también en infarto al miocardio), se recomienda seleccionar una escala razonablemente fiable, válida y sensible al cambio de cualesquiera de los instrumentos, ya sea genérica y específica. Asimismo, y tomando en cuenta el entorno en México, en vista de que la mayoría de los instrumentos han sido diseñados en idiomas diferentes al español, conviene explorar si la escala seleccionada tiene una versión al castellano, ya que se ha comprobado que los conceptos de la calidad de vida varían por el idioma. Además, es conveniente tomar en cuenta el entorno social de las poblaciones donde fue diseñada la escala, porque la calidad de vida puede estar influenciada por asuntos económicos, culturales o educativos; esto se pone de manifiesto cuando se trata de implementar un instrumento en países pobres, cuando fue concebido en países desarrollados, ya que las condiciones sociales hacen que los individuos perciban de manera muy diferente un síntoma o una discapacidad. Por lo anterior, sólo los instrumentos que hayan sido traducidos y adaptados a las poblaciones específicas deberían ser los que se seleccionen para su posible aplicación en la clínica o en investigación. La traducción y adaptación de instrumentos es una rama en la investigación de Epidemiología Clínica.

## Instrumentos para medir calidad de vida relacionada con la salud

### Instrumentos genéricos

Se han desarrollado para ser empleados en diferentes tipos de pacientes o poblaciones; en general, la calificación de la calidad de vida que se desprende de estos instrumentos

es independiente de la enfermedad que afecta a los pacientes, por lo que permiten un nivel de comparación más global y abstracto sobre el daño que producen las diferentes enfermedades. Estos instrumentos además son útiles para explorar la eficiencia con que se asignan los recursos, por lo que pueden ser de ayuda en la planeación sanitaria; se les puede dividir en tres subgrupos:

1. **Medidas de pregunta única.** Consisten en hacer preguntas al paciente acerca de su salud, siendo el siguiente tipo de pregunta el más utilizado: “¿cómo diría que se encuentra de salud en este momento?” El paciente responde en una escala ordinal, que va de muy bien a muy mal, para lo cual es común que se utilicen escalas análogas-visuales.
2. **Perfiles de salud.** Es el genérico más común, mide diferentes dimensiones de la calidad de vida relacionada con la salud. La principal ventaja es que tratan de obtener un perfil general sobre la opinión percibida acerca de la calidad de vida de los diferentes individuos o grupos poblacionales mediante la valoración indirecta de cuestionarios personales. La principal limitación es que, al ser instrumentos generales, pueden no adecuarse a los aspectos de la calidad de vida más importantes de la enfermedad concreta que se estudia y, por tanto, resultar un instrumento poco sensible a los cambios producidos por el tratamiento o la aparición de algunas complicaciones.

Los perfiles de salud más utilizados son: el Perfil de Salud de Nottingham (*Nottingham Health Profile*, NHP), el Cuestionario de Salud SF-36 (*Short Form-36 Health Survey*, SF-36), el Perfil de las Consecuencias de la Enfermedad (*Sickness Impact Profile*, SIP), el Cuestionario de Evaluación Funcional Multidimensional OARS (*OARS Multidimensional Functional Assessment Questionnaire*), entre otros muchos.

3. **Medidas de utilidad o preferencia.** En general, son los instrumentos que se utilizan en los estudios o análisis económicos. Se basan en la determinación de las utilidades o preferencias que los individuos asignan a los diferentes estados de salud del instrumento, de forma que proporcionan una puntuación única que refleja numéricamente la calidad de vida. Por lo general, se sitúan en una escala que va de 0 (peor estado de salud imaginable, a veces la muerte) a 1 (mejor estado de salud imaginable). Son medidas que se aproximan a suministrar un valor que refleja la calidad de vida, lo cual se considera apropiado para calcular los años de vida ajustados por calidad (en inglés, QALY [*Quality Adjusted Life Years*]) y los años de vida ajustados por discapacidad (DALY [*Disability Adjusted Life Years*]), los cuales son medidas de los estudios de costo-efectividad, costo-utilidad y costo-beneficio.

Los instrumentos más utilizados son: Euroqol 5-D, *Quality of Well Being Scale* y la Matriz de Rosser y Kind.

## Instrumentos específicos o funcionales

Se han desarrollado para utilizarse en pacientes con una enfermedad concreta (asma, diabetes, depresión, etc.) y, por tanto, pueden detectar mejor cómo un problema de

salud en particular afecta a la calidad de vida de este tipo de pacientes y son sensibles a variaciones de la enfermedad o de las intervenciones. La dimensión fundamental a la que se dirige es a medir la eficacia, efectividad y eficiencia de los procedimientos diagnósticos y terapéuticos de una enfermedad en particular, por lo que la utilidad central va encaminada a la evaluación de tecnologías sanitarias desde el punto de vista clínico y/o económico.

La principal ventaja es que presentan una alta sensibilidad a los cambios ante un problema de salud, especialmente indicados en los estudios que pretenden medir el cambio en la calidad de vida tras una intervención. La principal limitación es que no permiten realizar comparaciones entre las diferentes enfermedades. Son múltiples los ejemplos, principalmente en relación con enfermedades crónicas, de los cuales existen algunos adaptados para su uso en español.

Por último, la decisión de seleccionar entre un instrumento genérico o específico para medir la calidad de vida depende de los objetivos que se persigan; así, los genéricos son especialmente útiles para comparar la calidad de vida entre diferentes poblaciones y padecimientos, por lo que se pueden considerar meramente descriptivos; mientras que los instrumentos específicos están diseñados para detectar si los cambios físicos o efectos del tratamiento modifican la calidad de vida en pacientes con una enfermedad o condición en particular.

## Cuestionario de calidad de vida SF-36

El cuestionario SF-36 (*Short Form 36 Health Survey*) es un instrumento o escala que se ha utilizado en investigaciones clínicas desde principios de la década de 1990 para determinar la calidad de vida de los individuos; es uno de los instrumentos genéricos más usados y tiene traducciones validadas en diferentes idiomas y países, incluyendo a México. El SF-36 ofrece una perspectiva general del estado de salud de la persona, con la ventaja de que es autoaplicable, fácil y rápido de llenar, a la vez que también es sencillo de evaluar.

El SF-36 permite evaluar numéricamente diferentes aspectos de la salud de la persona. Todas las dimensiones son calificadas en una escala que varía de 0 a 100, siendo el valor 0 el peor estado de salud y el 100 el que representa el mejor estado. Está integrado por 36 preguntas que abordan diferentes aspectos relacionados con la vida cotidiana de las personas. Las preguntas se agrupan y miden de manera independiente en ocho dimensiones, las cuales comprenden dos aspectos fundamentales: la salud física y la salud mental (cuadro 15-1).

La figura 15-1 muestra un ejemplo de los reactivos que componen la escala SF-36.

Otra ventaja del SF-36 es que, en la actualidad, se tiene disponible en Internet una herramienta (<http://www.sf-36.org/nbscalc/index.shtml>) que hace fácil su cálculo e interpretación, comparando la puntuación obtenida de un paciente con la de la población general, de acuerdo a cada una de las dimensiones.



**Cuadro 15-1.** Dimensiones del cuestionario SF-36 para medir la calidad de vida, de acuerdo con sus dos componentes principales

| Salud física                                                               | Salud mental                                                                  |
|----------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| <b>1) Funcionamiento físico</b><br>(PF – <i>physical functioning</i> )     | <b>1) Funcionamiento o rol social</b><br>(SF – <i>social functioning</i> )    |
| <b>2) Limitación por problemas físicos</b><br>(RP – <i>role physical</i> ) | <b>2) Salud mental</b><br>(MH – <i>mental health</i> )                        |
| <b>3) Dolor corporal</b><br>(BP – <i>bodily pain</i> )                     | <b>3) Limitación por problemas emocionales</b><br>(RE <i>role emotional</i> ) |
| <b>4) Percepción general de la salud</b><br>(GH – <i>general health</i> )  | <b>4) Vitalidad, energía o fatiga</b><br>(VT – <i>vitality</i> )              |

## Resolución del escenario inicial

El médico internista decidió utilizar para la paciente con artritis reumatoide el cuestionario SF-36 para evaluar la calidad de vida. Después de hacer la evaluación a la paciente, el médico ingresó en el sitio en Internet la información obtenida. El resultado se muestra en la figura 15-2.

Como se observa, antes del inicio de rituximab la paciente tenía, en general, calificaciones más bajas de su calidad de vida en comparación con las registradas de una población estándar. En la gráfica de la izquierda se muestran las calificaciones de las ocho dimensiones; con excepción de la percepción general de la salud (GH) y de la vitalidad (VT), el resto se encuentra por debajo del promedio obtenido de la población general (línea continua por arriba de las barras). Esto mismo se observa en la gráfica del centro,

### Ejemplo de reactivos del SF-36

| 3. Las siguientes preguntas se refieren a actividades o cosas que usted podría hacer en un día normal. Su salud actual, ¿le limita para hacer esas actividades o cosas? Si es así, ¿cuánto? |                            |                            |                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|----------------------------|----------------------------|
|                                                                                                                                                                                             | Sí, me limita mucho        | Sí, me limita un poco      | No, no me limita nada      |
| <b>A) Esfuerzos intensos</b> , tales como correr, levantar objetos pesados, o participar en deportes agotadores                                                                             | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| <b>B) Esfuerzos moderados</b> , como mover una mesa, pasar la aspiradora, jugar a los bolos o caminar más de 1 hora                                                                         | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| <b>C) Coger o llevar la bolsa de la compra</b>                                                                                                                                              | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| <b>D) Subir varios pisos</b> por la escalera                                                                                                                                                | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| <b>E) Subir un solo piso</b> por la escalera                                                                                                                                                | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |

**Figura 15-1.** Ejemplo de reactivos del SF-36.



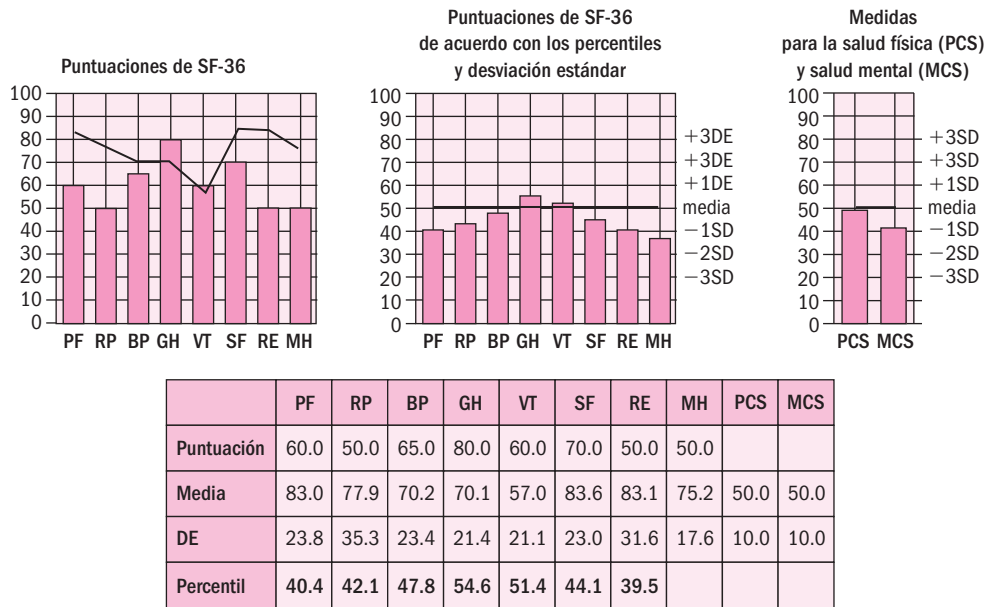


**Figura 15-2.** Puntuación de SF-36 de paciente con artritis reumatoide en comparación con una muestra poblacional estándar (EUA, 1998).

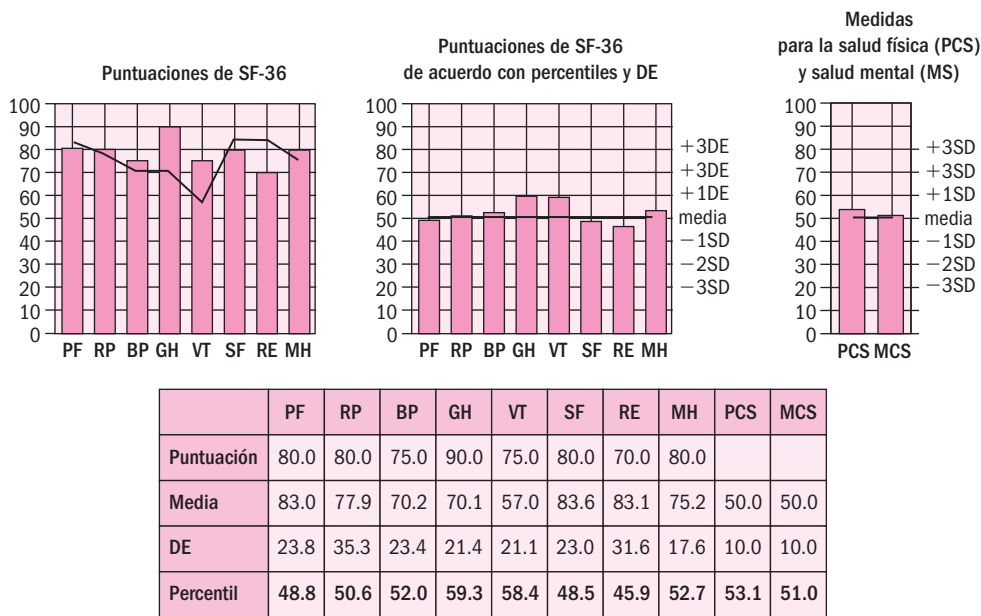


**Figura 15-3.** Puntuación de SF-36 de paciente con artritis reumatoide a seis meses de tratamiento con rituximab.

pero de acuerdo con los valores promedio, mientras que en la gráfica de la derecha se integran los dos componentes principales, donde la salud física se encuentra dentro del promedio de la población, pero no la salud mental, que está por debajo de una desviación estándar. En la tabla inferior se despliegan los valores puntuales, tanto el de la paciente como el de la población de referencia.

Después de seis meses de tratamiento con rituximab, la paciente presentó mejoría desde el punto de vista clínico, ya que hubo disminución de la actividad de la artritis. Su médico le solicitó volver a evaluar su calidad de vida utilizando el mismo sistema; en la figura 15-3 se presentan los resultados.

Como es evidente, la paciente mostró mejoría en la calidad de vida en las ocho dimensiones, incluso en aquellas donde estaba por debajo de los valores de referencia. De esta manera, el médico interpretó que la mejoría clínica se relacionó con mejoría en la calidad de vida, la cual puede ser comparable con la de la población de referencia.

## Bibliografía

- Alonso J, Antó JM, Moreno C. Spanish version of the Nottingham Health Profile: Translation and preliminary validity. *Am J Public Health* 80:704-708. 1990.
- Alonso J, Prieto L, Antó JM. La versión española del SF-36 Health Survey (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. *Med Clin (Barc)* 104:771-776. 1995.
- Anderson JP, Bush JW, Berry CC. Classifying function for health outcome and quality of life evaluation. *Med Care* 24:459-469. 1986.
- Arias-Gómez J, Hernández-Hernández D, Benítez-Aranda H, Villasís-Keever MA, Bernáldez-Ríos R, Martínez-García MC. Un instrumento para medir la calidad de vida por medio del desempeño diario en pacientes pediátricos con leucemia. *Gac Med Mex* 132:19-28. 1996.
- Arias-Gómez J, Salvador-Garduño N, López-García A, Cárdenas-Navarrete R, Serret-Montoya J, Villasís-Keever MA. Concepto de calidad de vida en el adolescente sano y enfermo. *Bol Med Hosp Infant Mex* 55:707-711. 1998.
- Arostegi Barandika I. *Evaluación de la calidad de vida en personas adultas con retraso mental en la Comunidad Autónoma del País Vasco*. Tesis doctoral. Consultado diciembre del 2009: <http://campus.usal.es/~inico/investigacion/jornadas/jornada3/actas/simp29.pdf>
- Badia Llach X, Lizán Tudela L. Estudios de calidad de vida. En: Martín-Zurro A, Cano Pérez JF. *Atención primaria: conceptos, organización y práctica clínica*, 5a. ed. Elsevier, Madrid, España, pp. 251-259, 2003.
- Badia X, Roset M, Montserrat S, Herdman M, Segura A. La versión española del EuroQol: descripción y aplicaciones. *Med Clin (Barc)* 112 (suppl. 1):79-86. 1999.
- Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a Health Status Measure. *Med Care* 19:787-805. 1981.
- Centers for Disease Control and Prevention. *Measuring Healthy Days. Population assessment of health-related quality of life*. CDC, Atlanta, Georgia, EUA, November 2000.
- Fillenbaum GG, Smyer MA. The development, validity and reliability of the OARS. Multidimensional functional assessment questionnaire. *J Gerontol* 36:428-434. 1981.

- García Gómez JL. Evaluación de la calidad de vida en los pacientes con diabetes mellitus tipo 1: el caso del Hospital Universitario Ntra. Sra. de Candelaria de Tenerife. Tesis doctoral. Consultado en diciembre de 2009: <ftp://tesis.bbt.ull.es/ccssyhum/cs213.pdf>
- García Martínez A, Saez Caneras J, Escarbajal de Haro A. Educación para la salud: conceptos y definiciones. En: *Educación para la salud: La apuesta por la calidad de vida*. ARAN Ediciones, Madrid, España, pág. 260. 2000.
- González de Dios J. Calidad de vida relacionada con la salud: conocer e implementar en la toma de decisiones basada en pruebas en pediatría. *An Pediatr (Barc)* 60:507-513. 2004.
- López Izuel Ch, Rui Camps M, Forner Bscheid M. Enfermedad y calidad de vida, capítulo 2. En: *Cuidados enfermeros*. Masson, Barcelona, España, pp. 19-22. 2004.
- López-García A, Valois L, Arias-Gómez J, Alonzo-Vázquez F, Cárdenas-Navarrete R, Villasís-Keever MA, Martínez-García MC, Morales-Castillo ME. Validación del cuestionario COOP-Daurtmouth para evaluar estado funcional biopsicosocial en escolares y adolescentes con enfermedad crónica. *Bol Med Hosp Infant Mex* 53:606-615. 1996.
- Martínez Vizcaíno V. Introducción. *Calidad de vida en los ancianos*. Ediciones de la Universidad de Castilla-La Mancha, España, pp. 13-15. 1998.
- Pane S, Solans M, Gaite L, Serra Sutton V, Estrada MD, Rajmil L. *Instrumentos de calidad de vida relacionada con la salud pediátrica: Revisión sistemática de la literatura: actualización*. Agencia de Evaluación de Tecnología e Investigación Médicas. Barcelona, España, 2006.
- Patrick DL, Erickson P. *Health status and health policy. Allocation resources to health care*. Oxford University Press, Nueva York, EUA, 1993.
- Pope AM, Tarlov AR (eds.). *Disability in America: Toward a National Agenda for Prevention*. Committee on a National Agenda for the Prevention of Disabilities, Institute of Medicine, 1991.
- Rapley M. Introduction. Where has QOL come from? En: Rapley M. *Quality of life. Research. A critical introduction*. SAGE Publications Inc., California, EUA, pp. 4-25. 2003.
- Reig-Pintado E, Garduño-Estrada L. Procesos cognoscitivos asociados a los constructos de calidad de vida y bienestar subjetivo. En: Estrada Garduño L, Salinas Amescua B, Rojas Herrera M. *Calidad de vida y bienestar subjetivo en México*. Universidad de las Américas, Puebla. Plaza y Valdés, México. 2005.
- Rosser R, Kind P. A scale of valuations of states on illness: Is there a social consensus? *Int J Epidemiol* 7:347-358. 1978.
- Schwartzmann L. Calidad de vida relacionada con la salud: aspectos conceptuales. *Ciencia y Enfermería* 9:9-21. 2003.
- Velarde Jurado E, Ávila Figueroa C. Evaluación de la calidad de vida. *Sal Pub Mex* 44:349- 361. 2002.
- Zúñiga MA, Carrillo-Jiménez GT, Fos PJ, Gandek B, Medina-Moreno MR. Evaluación del estado de salud con la Encuesta SF-36: resultados preliminares en México. *Salud Públ Mex* 41:110-118. 1999.

# Capítulo 16

## Evaluación de la calidad de la atención médica

*Juan Garduño Espinosa*



### Introducción

La calidad de la atención médica deriva de preguntas básicas inherentes a la naturaleza misma de la Medicina. Desde una perspectiva social, algunas de estas interrogantes resultan fundamentales para la justificación de la práctica médica, por ejemplo: ¿en qué grado la atención en materia de salud es útil para la población que la recibe?, ¿en qué medida la atención médica en su conjunto hace más bien que mal?, ¿cuál es el grado de calidad que prestan los servicios de salud en México? Durante las últimas décadas, la calidad de la atención ha alcanzado un desarrollo tal que a la fecha puede describirse y cuantificarse, aun cuando los resultados obtenidos con dicha evaluación deban interpretarse con cautela, cuando menos en algunos casos.

La importancia de este aspecto de la práctica médica ha sido resaltada por Arnold Relman, editor de la revista médica *New England Journal of Medicine*, quien ha denominado a este esfuerzo por evaluar la calidad de la atención, **una tercera revolución en la atención médica**.

Parte de la preocupación por evaluar la calidad de la atención se relaciona con los crecientes costos que se derivan de proporcionar servicios, los cuales, a su vez, no necesariamente impactan en una mejoría clara en el estado de salud de la población. En consecuencia, por ejemplo, en 1989 se gastaron 600 mil millones de dólares en Estados Unidos para la atención de la salud, lo que representó 11% del producto interno bruto de ese país; esta cifra significa 2.8 veces más de lo que se invirtió en la Gran Bretaña. Aparentemente, sin embargo, los resultados en cuanto a la salud de la población en ambas naciones son similares.

Por otra parte, la tasa de mortalidad infantil y la esperanza de vida en Estados Unidos son iguales a las que se observan en países con un gasto mucho menor en el área de salud, como es el caso de Australia, Austria, Italia, Noruega, Suecia y Holanda.

Se han descrito dos acepciones de la palabra **calidad** relacionadas con el tema que aquí se trata: por una parte, “calidad” es una característica o propiedad de un objeto y, por otra, se refiere a tener superioridad de cierto tipo. Derivado de esto, el término enfatiza dos cosas, ya sea la preferencia por un objeto o el valor que se le asigna al mismo. Se ha establecido una cierta discusión sobre si la calidad debe considerarse en el sentido metafísico, ya mencionado (Donabedian afirma que la calidad es una propiedad que puede existir en grados variables), o bien si ha de dársele un sentido preferencial, como se describe en la segunda acepción, es decir, la del valor.

A partir de un enfoque clínico, sin embargo, los dos principales objetivos al evaluar la calidad de la atención médica son, por una parte, disminuir al mínimo posible la ocurrencia de efectos adversos en los pacientes y, por la otra, ofrecer a los enfermos la mejor experiencia posible, ya sea del médico o del Sistema de Salud, cuando surge la necesidad de recurrir a la atención médica.

En general, son tres las estrategias actualmente en uso para evaluar la calidad de atención a través del estudio de la competencia clínica de los médicos. La primera de ellas se efectúa mediante la medición de la participación en cursos de educación médica continua; en segundo lugar, por medio de los procesos de certificación de los médicos en las organizaciones que agrupan a las diferentes especialidades clínicas, y la tercera, mediante la evaluación del desempeño de los clínicos a través de auditorías clínicas, en las cuales se evalúa lo que cada médico hace en cada paciente, comparando su accionar con estándares establecidos para determinada enfermedad.

En la búsqueda de proporcionar la mejor calidad de atención a la población, es necesario identificar, desde un inicio, cuáles son los estándares de calidad que se aspira a cumplir en condiciones ideales de la práctica (eficacia de la atención); además se han de desarrollar mecanismos que permitan medir si dichos estándares se logran en la práctica médica cotidiana (evaluación de la calidad de la atención o efectividad de la misma), al menor costo posible (eficiencia en la atención) y, finalmente, se debe pretender desarrollar mecanismos que permitan solucionar las deficiencias identificadas (garantía de calidad).

## Definición de calidad de la atención

No se cuenta con una definición de **calidad de la atención médica** que sea satisfactoria para todos. Existen varios problemas que explican esta situación, de los cuales deben mencionarse, en primer término, los relacionados con eficacia y equidad. Otros problemas son la consideración de los costos y las preferencias y expectativas del paciente. En relación con el problema de la eficacia, se debe reconocer que el grado de conocimientos de que se dispone para atender a los sujetos sanos y enfermos es todavía muy limitado. Los avan-

ces científicos en la Medicina están muy lejos de permitir su aplicación en forma amplia y con la precisión requerida a los pacientes individuales.

La escasa información sobre calidad de la atención, obtenida científicamente y de la que actualmente disponemos, tiene, en general, problemas importantes de validez (debido a lo imperfecto de los métodos para generarla). En último término, esta información sólo podría aplicarse a pacientes promedio en condiciones ideales, ya que no toma en cuenta la enorme amplitud de variaciones individuales de los sujetos, así como de los sistemas de salud que los atienden.

En consecuencia, definir cuál es la mejor atención médica es difícil, dado lo incompleto e imperfecto de la información científica disponible. La respuesta dependerá de la especificidad de las preguntas: ¿en qué contexto clínico?, ¿con qué nivel de recursos disponibles? Por otra parte, aquellos aspectos de la práctica médica que se sustentan sobre la experiencia acumulada por la profesión a lo largo de los siglos están sujetos a variaciones en el juicio que los profesionales hacen acerca de los mismos. De esta forma, establecer cuándo un paciente ha sido atendido con calidad constituye un asunto que no puede calificarse con total certidumbre.

La ponderación de cada uno de los aspectos involucrados en la atención proporcionada está también abierta a discusión, para lo cual no será fácil tampoco alcanzar un consenso. Aunado a estas dificultades, se encuentra la consideración a la siguiente pregunta: ¿quién debe juzgar la calidad de la atención prestada? Parecería obvio que debiera ser el personal de salud, con la autoridad moral que se deriva de ser parte de la comunidad médica (si se considera la información científica válida disponible hasta tal momento y la mejor experiencia acumulada). El resultado sería que, desde el punto de vista de la beneficencia médica, podrían identificarse los aspectos que determinan (en forma jerárquicamente ordenada) la mejor atención médica disponible en dicho momento.

Sin embargo, al respecto ha de reconocerse el problema de que si sólo se considera al personal de salud, se excluyen las preferencias de los enfermos acerca de los resultados que se consideran deseables al recibir atención médica. Es claro, por lo demás, que la evolución social ha conducido a un mayor énfasis en la autonomía de los pacientes, lo que ha condicionado también un decremento en la importancia del principio de beneficencia médica.

Por otra parte, la disponibilidad de recursos no es homogénea, lo cual conduce al problema de la equidad. Casi siempre los expertos médicos se encuentran en donde se dispone de todos los recursos y con base en esta amplia disposición, pueden generarse normas ideales de atención. En consecuencia, no es posible aplicar las mismas recomendaciones en todas partes. Esta situación causa el surgimiento de diversas interrogantes cuya respuesta no se halla con facilidad, como “¿existen diferentes niveles de calidad de acuerdo con la calidad con que se cuente para tener la tecnología necesaria?”; “¿quién fija los distintos niveles de calidad y según qué criterios?”; “desde una perspectiva social, ¿es aceptable, por parte de las comunidades médicas, proporcionar diferentes niveles de calidad a los pacientes en función de su capacidad financiera individual?”

Como puede verse, la dificultad para definir el ámbito de la calidad de la atención médica representa un reto de primer orden en los aspectos intelectual, moral y operacio-

nal para todo el personal de salud. Sin embargo, a pesar de lo complejo de la definición del término, cabe analizar algunas de las definiciones propuestas.

El Instituto de Medicina ha definido la calidad de la atención como el grado en que los servicios de salud para los individuos y las poblaciones aumentan la probabilidad de obtener resultados de salud deseables y que sean consistentes con el nivel de conocimientos profesionales vigentes en ese momento. Esta definición, también propuesta por el Comité de Estudios del Programa de Garantía de Calidad para Medicare, es similar a las definiciones de la *Joint Commission on Accreditation of Health Care Organizations* y por la *Office of Technology Assessment* de Estados Unidos.

Un aspecto en el que esta definición hace énfasis es sobre la obtención de un resultado que sea deseable para los pacientes, a partir de lo cual se asume que éstos serán informados y participarán con sus médicos en la toma de decisiones acerca de su atención.

Wentzel define la calidad en función de tres componentes de la práctica médica, que son efectividad, eficiencia y aceptabilidad. La efectividad se refiere a la mejoría en el estado de salud que produce la atención médica proporcionada; la eficiencia señala el grado en que se alcanzan resultados con el menor número de recursos disponibles, y la aceptabilidad es el grado en el que las medidas aplicadas son congruentes con las expectativas del paciente.

La Asociación Médica Estadounidense (AMA), en 1984, definió la “calidad de la atención” como aquella que de modo consistente contribuye a mejorar o mantener la calidad o duración de la vida. En 1986, esta asociación agregó ocho elementos inherentes al proceso de atención, que sumados al resultado obtenido por el paciente, definen la calidad de la atención que se proporciona. Los ocho elementos son: 1) la atención médica produce una mejoría óptima en la salud del individuo; 2) se enfatiza la promoción de la salud y la prevención de la enfermedad; 3) se proporciona oportunamente; 4) busca obtener la participación y cooperación informada del paciente, en el proceso de atención y en las decisiones que se toman en relación con su salud; 5) está basada en los conocimientos derivados de la ciencia médica; 6) se proporciona con sensibilidad y preocupación por el bienestar del enfermo; 7) utiliza en forma eficiente a la tecnología, y 8) se encuentra lo bastante documentada para dar continuidad a la atención y permitir que sea posible evaluarla. En el marco de esta definición, cuando se cumplen estos ocho elementos y se obtiene un resultado favorable por parte del paciente, puede argumentarse una elevada calidad en la atención proporcionada.

De acuerdo con Donabedian, la calidad es la magnitud con la que se espera se obtenga la atención proporcionada, para el balance más favorable entre riesgos y beneficios. En otro momento, a esta definición se agregó el concepto de costos, en donde los beneficios menos los riesgos y costos constituyen lo que se ha denominado **modelo unificador de calidad de la atención**.

Rutstein define a la calidad como el efecto de la atención sobre la salud del individuo, la cual debe diferenciarse de la eficiencia de la atención médica. Desde el punto de vista de dicho autor, la **calidad** se refiere al resultado obtenido por el paciente y la **eficiencia** está relacionada con el proceso de atención. Para Rutstein, la forma en que



puede evaluarse la calidad de la atención es mediante la identificación de efectos adversos en el enfermo, por ejemplo, procesos de mortalidad o complicaciones. A partir de esta identificación, es posible proponer mejoras al sistema al buscar entre el proceso de atención las causas que presumiblemente dieron origen al mal resultado.

Una limitante de este enfoque es la necesidad de ajustar los resultados obtenidos por el paciente con otras variables, en especial la gravedad de la enfermedad con la que el sujeto acude a solicitar atención médica. Por otra parte, debe tomarse en cuenta que muchas complicaciones ocurren después de que el individuo ha egresado del hospital y que indebidamente podrían no tomarse en cuenta al momento de la evaluación. En ocasiones, los hospitales que comunican los peores resultados son aquellos que cuentan con los mejores sistemas de vigilancia para dichos procesos adversos, lo cual ha sido documentado en el caso de las infecciones respiratorias.

## Modelos de evaluación de la calidad de la atención

Un modelo de calidad de la atención médica que ha resultado de particular importancia para el desarrollo de esta área, ha sido el propuesto por Donabedian, el cual divide la atención médica en tres componentes: estructura, proceso y resultado.

### Estructura

Se refiere al conjunto de recursos físicos, materiales y de personal de salud con que se cuenta para brindar la atención a la salud. Incluye las clínicas y los hospitales, el número de camas disponibles, los recursos tecnológicos disponibles para el diagnóstico y tratamiento de los pacientes, los medicamentos, el número y tipo del personal de salud y aun los grados académicos de dicho personal. La Comisión de Acreditación de Hospitales de Estados Unidos, por ejemplo, evalúa algunos elementos de estructura en un hospital, tales como la existencia de equipo de monitoreo en los quirófanos y que se programen conferencias de mortalidad y morbilidad relacionadas con la revisión de complicaciones quirúrgicas indeseables.

### Proceso

Se refiere a todos los procedimientos diagnósticos y terapéuticos que se realizan con objeto de resolver el problema del paciente. Incluye el hecho de que los expedientes clínicos se encuentren firmados y con notas acerca del enfermo registradas de forma oportuna, uso apropiado de exámenes de laboratorio y gabinete, así como que los aspectos relevantes del examen físico se hayan realizado y registrado. Los estándares de calidad de la atención médica deben especificarse tanto para el proceso como para el resultado. Es decir, con las medidas actualmente disponibles del proceso, es necesario hacer explícito cuáles son los mejores resultados para el paciente que se debe aspirar a cumplir, por ejemplo, ¿cuál es el tiempo promedio de sobrevida que debe alcanzar un paciente con cáncer de próstata?, ¿qué grado de calidad de vida ha de esperarse obtener en un enfermo con insuficiencia renal?



Los estándares tienen que referirse tanto a aspectos técnicos como biológicos y psicológicos. Esto es válido para las medidas del proceso y para las de resultado; sin embargo, hasta la fecha, la mayoría de los estándares de calidad desarrollados se refieren a aspectos técnicos del proceso. La dificultad inherente a medir los aspectos psicológicos (p. ej., los de la relación médico-paciente) ha retrasado el avance en esta parte fundamental del proceso.

Un buen proceso de atención debe demostrar sus beneficios a través de la realización de estudios clínicos de eficacia y con frecuencia es deseable que muestre su grado de costo y efectividad. Algunas dificultades añadidas al desarrollo de instrumentos que midan los aspectos técnicos de la calidad de la atención, se derivan de que si bien es posible medir con relativa facilidad si un estándar se cumple, es muy difícil medir qué tan bien se ha cumplido. Un problema agregado es el hecho de que la evaluación del proceso, frecuentemente, requiere obtener la información a través de los expedientes clínicos, los cuales pueden haber desaparecido o estar incompletos y a veces incluso es posible que resulten ilegibles.

## Resultado

Alude a la forma en que respondió el paciente a la enfermedad y a la atención recibida. Los resultados que con más frecuencia se consideran son si el individuo sobrevive o muere, o bien las complicaciones e incapacidades que ocurrieron al enfermo. Esta medición ha de incluir, además de los resultados obtenidos en cuanto a morbilidad y mortalidad, aquellos relativos a la disminución de los síntomas y mejoría en las funciones cotidianas, o en la sensación de bienestar y en la calidad de vida relacionada con la salud, así como en el grado de satisfacción de los pacientes. Estos aspectos, denominados **características psicológicas del resultado**, en general incluyen la dinámica propia de los sucesos ocurridos al enfermo; el grado de satisfacción con la atención recibida; el grado de cumplimiento de las indicaciones, y el estado de salud después de recibir la atención.

Los primeros estudios de calidad de la atención se enfocaron en aspectos de estructura; posteriormente se identificó la importancia de incorporar elementos del proceso en esta evaluación y, en los últimos años, se ha concedido importancia prioritaria al resultado que obtienen los enfermos a consecuencia de la atención que reciben.

Los estándares referentes a un buen resultado deben alcanzarse a través de estudios clínicos rigurosamente diseñados, para evaluar el curso clínico de los pacientes tratados con los mejores recursos disponibles. Sin embargo, a la fecha, la mayoría de los estándares, tanto de proceso como de resultado, no han sido probados por medio de investigación científica rigurosa. En general, aquellos aspectos de proceso considerados como de los más aceptables, en su mayoría se han evaluado empíricamente, es decir, descansan en su mayor parte en la experiencia individual y colectiva.

El proceso de atención médica se ha dividido en dos aspectos: técnico e interpersonal. El primero resulta de la aplicación de la ciencia y la tecnología de la Medicina, así como de otras ciencias relacionadas con la salud, al manejo de un problema individual de salud, de manera tal que se obtenga el máximo beneficio con los menores riesgos.

El aspecto interpersonal se produce por la interacción social y económica entre el médico y su paciente, al considerar valores, normas y expectativas de ambos, así como la ética de la profesión médica. Relacionadas con el proceso de atención se encuentran las denominadas **amenidades**, las cuales se refieren a las características de lugar y equipo en donde se presta la atención médica y que pueden describirse en términos de comodidad, privacidad, cortesía, aceptabilidad, entre otras.

## Mala calidad de la atención

El concepto de **calidad de la atención médica** o **auditoría clínica** lleva implícitos dos aspectos. Por una parte, implica cuantificar la práctica médica en términos de los procedimientos realizados y los resultados obtenidos en el enfermo; por la otra, conlleva la necesidad de evaluar lo realizado, es decir, ejercer algún tipo de control de dichas prácticas y procedimientos, esto último con objeto de detectar niveles subóptimos de rendimiento en relación con estándares establecidos para que, en el caso de identificarse, pueda proponerse algún remedio.

Los pacientes esperan que los médicos que los atienden sean competentes. Una forma de definir la competencia es en relación con algo apropiadamente calificado. En la medida en que el personal es efectivo y eficiente, contribuye a la calidad de la atención que se presta a los individuos. Se debe generar evidencia a través de la investigación médica acerca de problemas en la calidad prestada por el Sistema de Salud, así como en relación con la frecuencia con la que ocurre y los factores que se asocian con la misma.

La mala calidad de la atención médica se ha clasificado de acuerdo con la sobreutilización de los servicios de salud, infrauso o pobre rendimiento técnico o interpersonal. La evidencia de sobreuso, especialmente en lo que se refiere a algunos procedimientos y medicamentos, es amplia. Así, en una revisión de 5 000 expedientes, 17% de los procedimientos de angiografía coronaria y 32% de las endarterectomías se juzgaron como inapropiadas y se consideró que 9 y 32% de estos procedimientos, respectivamente, se habían efectuado por indicaciones dudosas. La subutilización de algunos servicios se ha documentado en ciertos grupos de edad, particularmente en ancianos; por ejemplo, se ha informado una frecuencia subestimada de diagnósticos en casos de incontinencia urinaria, infecciones, trastornos metabólicos y problemas psiquiátricos (en especial, depresión) en este grupo de edad. En cuanto al uso disminuido de servicios, se ha observado subutilización de tratamiento con beta bloqueadores, en pacientes ancianos que han padecido infarto agudo al miocardio.

Los antibióticos a menudo se utilizan de modo inadecuado. Por ejemplo, en un estudio pudo identificarse que sólo en 72% de los casos en que fueron indicados para fines terapéuticos y en 36% de aquellos en que se administraron con fines profilácticos, su uso pudo considerarse como apropiado. En algunos lugares se ha estimado que entre 6 y 8% de los médicos tienen problemas serios y recurrentes en lo que se refiere al grado de calidad que proporcionan. Se ha documentado que la tasa de práctica médica inadecuada o daños a la salud derivados de ésta, es de aproximadamente 4 a 5% por cada 100

hospitalizaciones y se calculó que 17% de la misma se debe a negligencia. A principios del decenio de 1980-1989 se habían identificado más de 1 000 estudios que documentaron deficiencias en la calidad de la atención proporcionada.

En personas con síndrome gripal, entre 20 y 70% reciben atención inapropiada; en hospitales de enseñanza, la mitad de los antibióticos se utiliza en forma incorrecta y un tercio de todos los exámenes de laboratorio considerados como anormales, nunca se evalúa para decidir sobre el enfermo. En ocasiones se ha identificado que la calidad de la atención quizá no tenga sólo impacto sobre el paciente individual, sino incluso puede relacionarse con un potencial deterioro del ambiente. Así, por ejemplo, el uso de gases anestésicos, rayos X y sustancias radiactivas, elementos químicos utilizados en el laboratorio y algunos fármacos de tipo citotóxico, puede constituir un aspecto de la tecnología médica cuyo impacto sobre el medio está por determinarse.

## Criterios implícitos y explícitos

Se han descrito dos tipos de criterios para evaluar la calidad de la atención y se los ha definido como herramientas complejas, confiables y válidas que reflejan los atributos más relevantes de la estructura, el proceso y resultado, y que sirven para fines de comparación al momento de evaluar. Se han señalado dos tipos de criterios.

### Criterios implícitos

Consisten en la opinión del personal de salud experto en determinada área, respecto de la forma en que se ha llevado a cabo la atención prestada al paciente. La necesidad de utilizar juicios como este se basa en la complejidad de la práctica médica, la cual vuelve no viable el desarrollo de criterios para cada una de las fases de la atención. Este enfoque metodológico fue promovido desde 1958 por Butler y Quinlan, y ha sido utilizado en forma amplia en la revisión de la calidad de la atención en diversas áreas. Los criterios implícitos pueden ser estructurados o no estructurados. En el primer caso se solicita al revisor que califique la calidad prestada en diversos aspectos de la atención, con base en una escala preespecificada, si bien no se le pide que explique o defina los fundamentos en que se basa tal juicio.

Se ha certificado que, al utilizar este formato, las opiniones de los médicos permiten grados adecuados de reproducibilidad cuando se juzga la calidad que se ha prestado de manera global en un determinado lugar, por ejemplo: el grado en que podría haberse prevenido el que ocurrieran defunciones. Sin embargo, se ha mostrado también que el uso de criterios implícitos tiene bajos niveles de consistencia, cuando son utilizados por dos o más observadores al evaluar aspectos específicos de la calidad prestada; por ejemplo, el grado en que se usaron apropiadamente los recursos y la oportunidad con que los pacientes fueron dados de alta del hospital.

Los criterios implícitos no estructurados están dados por los juicios del personal de salud experto, con el propósito de juzgar la calidad de la atención médica, y en los cuales no se solicita justificación alguna para tales juicios. En esos casos es determinante estar

seguro de que el participante sea un experto. La utilización de este tipo de criterios ha sido un procedimiento muy empleado por el personal de salud, fundamentalmente entre los médicos, quienes los han utilizado con frecuencia para establecer juicios acerca de la calidad de su trabajo. Por ejemplo, la calidad de los expedientes clínicos se ha establecido tradicionalmente de acuerdo con la opinión, no basada en criterios objetivos, de los médicos con mayor grado de experiencia clínica.

## Criterios explícitos

Se refieren a la especificación previa de aquellos aspectos particulares que deben tomarse en cuenta para revisar el grado de calidad de la atención prestada en un área determinada. La utilización de este tipo de criterios puede facilitar que la revisión sea realizada por individuos no expertos en el área evaluada e incluso es posible que la efectúe personal no médico, el cual puede entrenarse debidamente en el uso de dichos criterios. Por ejemplo, estrategias clínicas para evaluar la calidad de la atención en que se utiliza este tipo de criterios son la formulación de protocolos para la atención, el desarrollo de guías clínicas y el establecimiento de reuniones de consenso.

Tres estrategias de desarrollo de estándares pueden identificarse para la evaluación de la calidad de la atención, como se muestra en el recuadro siguiente.

- **Guías de la práctica clínica.** Actualmente este aspecto ha tenido un amplio desarrollo, e intenta estandarizar la forma en que debe desarrollarse el proceso de atención médica en entidades específicas.
- **Mapas de criterios o árboles de decisión.** Son formas detalladas de descripción de la conducta médica ante un paciente individual. Toman en consideración todas las posibles complejidades en el diagnóstico y tratamiento de un enfermo y pretenden especificar todas las vías que podrían seguirse, para atender a todas las personas que acudan con un problema de salud particular.
- **Criterios de búsqueda de casos.** Estos criterios pretenden identificar pacientes en quienes haya ocurrido un problema en la atención médica y que requieran revisión profesional cuidadosa y más detallada en fecha posterior. Un ejemplo de esta técnica es la denominada **condición indicadora** o **método trazador**, cuyas características han especificado con detalle Kessner, Sibley y Chambers y que se describe más adelante.

Cada uno de estos enfoques metodológicos presenta diferentes ventajas y su utilidad depende del aspecto de la atención médica que se pretenda examinar.

## Tipos de estudios de calidad de la atención

Los estudios que se dirigen a evaluar la calidad de la atención pueden abordarse a partir de dos grandes enfoques, independientes entre sí, de acuerdo con la estructura del dise-

ño de investigación y según el nivel en que se realizan. En función de la estructura del diseño, es posible efectuar estas investigaciones a través de estudios descriptivos, ya sea longitudinales o transversales (encuesta descriptiva, cohorte descriptiva), comparativos observacionales (encuesta comparativa, casos y controles, cohortes comparativas) y experimentales. Los estándares de calidad de cada diseño de investigación son los mismos que aplican en cualquier otra área de la Medicina.

El segundo enfoque se refiere a dos posibles niveles en que puede llevarse a cabo la investigación: poblacional o clínico. En el primero, la información obtenida es necesaria para la toma de decisiones en términos de políticas de salud y se realiza fundamentalmente en el nivel social, es decir, aplica sobre todo en los sistemas de salud.

## Sistemas de salud

Cuando se cuenta con una gran cantidad de información a través de bases de datos, puede intentarse identificar indicadores relacionados con malos resultados en los pacientes. Estos indicadores también se denominan **eventos centinela**. Con este método, también es posible reconocer la existencia de servicios que se proporcionan en forma inapropiada. Otra técnica que puede utilizarse con este propósito es el análisis de las variaciones que ocurren en pequeñas áreas geográficas, con el fin de identificar diferencias en el uso de los servicios.

Las dificultades inherentes a la evaluación de la calidad de la atención en este nivel se explican en función de que los sistemas de salud proporcionan servicios, cuyo impacto es difícil de separar de otras maniobras sociales que contribuyen a la salud y al bienestar de la población.

Así, por ejemplo, se dice que en países desarrollados 97% de todos los niños sobrevive al periodo preescolar, en tanto que en naciones como México, de 20 a 25% de los niños muere antes de alcanzar esta fase de la vida.

¿En qué medida este resultado es un reto para el sistema de salud?, o los factores clave para la mejora ¿serán los relacionados con los sistemas de desarrollo social, como educación, alcantarillado, disponibilidad de agua potable, etcétera? Éste es un aspecto difícil de evaluar. Desviar la atención a los sistemas de salud, considerados como responsables de abatir problemas como el mencionado, puede generar una distracción para la sociedad en la búsqueda y el desarrollo de una solución al verdadero problema.

Una dificultad más al evaluar la calidad de la atención se deriva de la complejidad que se desprende de la aplicación de los elementos de juicio clínico a los sucesos evaluados. Esto es particularmente cierto cuando se pretende evaluar el desarrollo obtenido por los pacientes, por ejemplo, cuando se considera la mortalidad como la variable de resultado; esta evaluación es ineficaz por completo, si no se toma en cuenta la gravedad con la que ingresaron al hospital dichos pacientes.

Pollack identificó variaciones incluso seis veces mayores en la mortalidad observada entre seis hospitales pediátricos; sin embargo, estas diferencias desaparecieron cuando se ajustó la gravedad de la enfermedad con que habían ingresado los individuos. De la

misma forma, otras variables pueden ser fundamentales al momento de explicar un resultado; entre ellas se encuentran la comorbilidad de los pacientes, la edad, el servicio a través del cual ingresaron los pacientes, entre otras.

Knaus llevó a cabo un estudio en 13 unidades de cuidados intensivos y encontró diferencias incluso de tres veces entre ellas en lo que se refiere a tasas brutas de mortalidad; sin embargo, al ajustar de acuerdo con el diagnóstico indicaciones para el tratamiento y algunas mediciones fisiológicas, el resultado cambió de manera notable. En otras palabras, no considerar las características del enfermo al evaluar los resultados causa un incremento en la incertidumbre.

## Nivel clínico

Los hospitales pueden desarrollar indicadores, como las tendencias que siguen las infecciones hospitalarias o la existencia de otros procesos que se presentan con menor frecuencia, ya sea que esto ocurra a nivel de la unidad hospitalaria, de un servicio de la misma o de un médico en particular. Con este enfoque, es posible definir un indicador como toda situación clínica (enfermedad, signo, síntoma, etc.) que es razonablemente frecuente en la práctica médica y para la cual existen suficientes datos de que una buena atención médica produce beneficio al paciente. Algunos ejemplos de algunos indicadores desarrollados en diversos estudios son infección de vías urinarias, hipertensión arterial, otitis media, atención médica prenatal, cuidados del recién nacido, anemia, depresión, obesidad.

Otra técnica útil en el nivel clínico es la revisión de los expedientes médicos, ya sea por parte de los médicos o las enfermeras, donde se compara el proceso de atención con criterios implícitos, con el objetivo de juzgar la calidad en el proceso de atención médica. Para las dos técnicas descritas en el nivel clínico se debe asumir que los expedientes clínicos son de calidad aceptable.

Las investigaciones, tanto en el nivel poblacional como en el clínico, pueden considerarse indispensables; sin embargo, difieren en sus objetivos y en la precisión con que es posible documentar la información. El enlace entre ambas no es aún lo bastante claro y algunos esfuerzos se realizan en diversos lugares del mundo para construir un puente entre ellas.

## Relación proceso-resultado

Entre los estudios empíricos de calidad de la atención que se han realizado, ha generado particular preocupación identificar la falta de relación entre proceso y resultado. De confirmarse este hecho, podría ponerse en duda la utilidad de la práctica misma de la Medicina en algunas áreas; si los actos y procedimientos realizados por el personal de salud no tienen un mínimo de relación con la mejoría del paciente (es decir, una relación directamente proporcional: a mejor proceso de atención, mejores resultados en el enfermo), entonces la práctica médica pierde su razón de ser, por lo menos, en este aspecto.

Estudios realizados por Broock, Nobrega, Lindsay y Room han mostrado esta falta de relación. Dichos estudios, sin embargo, al ser analizados, permitieron identificar aspectos de metodología clínica, especialmente relevantes para la realización de estudios de calidad de la atención con un enfoque clínico. Un primer aspecto fue su falta de consideración al cumplimiento por parte del paciente, de las indicaciones dadas por el personal de salud. Se señala que esto podría contribuir a explicar la incapacidad para mostrar la relación mencionada; por ejemplo, otros estudios rigurosamente diseñados y que han tomado en cuenta el cumplimiento de los enfermos, han mostrado una asociación sólida entre proceso y resultado en lo que se refiere a hipertensión arterial, infecciones de vías urinarias y anemia por deficiencia de hierro.

Un elemento más de crítica a los estudios que han evidenciado falta de relación entre procesos clínicos y resultados en salud, es que no se ha tomado en cuenta el tamaño de la muestra necesario para identificar dicho grado de relación, considerado como clínicamente significativo. La duda persiste, a pesar de todo, si se considera que en estudios realizados en medios hospitalarios el cumplimiento terapéutico tiene menor importancia.

También se ha observado esta falta de relación entre proceso y resultado, en estudios relativos a cirugía electiva, apendicitis aguda e infarto agudo al miocardio. Parece fuera de duda que existe un espectro dentro de la práctica médica, en que algunas áreas muestran una relación casi perfecta entre el proceso y el resultado, y en el otro extremo se identifican procesos deficientes que pueden asociarse con buenos resultados.

Para interpretar debidamente la efectividad de la atención proporcionada por los servicios de salud y establecer de manera válida la relación entre procesos y resultados, es necesario considerar la gran heterogeneidad de pacientes que se incluyen en los estudios. En consecuencia, han de estudiarse todas aquellas características del individuo que aumentan el riesgo de un mal resultado, o bien, que influyen en la elección de un tratamiento. El propósito de considerar estas variables a fin de realizar un ajuste es buscar separar los efectos de la atención proporcionada, de aquellos elementos relacionados con el estado de salud basal o preexistente en el sujeto, así como discriminar el papel de otros factores como la edad o el estado socioeconómico, que podrían afectar también el resultado obtenido por el paciente.

A fin de cumplir con la condición ya mencionada, la medición del estado de salud preexistente debe incluir la evaluación de la gravedad de la enfermedad y de la comorbilidad (presencia de otras enfermedades coexistentes). Se ha sugerido, a fin de considerar lo anterior, incluir evaluaciones del estado funcional (capacidad de los enfermos para realizar sus actividades cotidianas) y del bienestar (evaluación global del estado de salud de los individuos), en el ajuste de casos cuando ingresan los pacientes al hospital, ya que estas medidas predicen la intensidad en el uso de los servicios y ha podido demostrarse su influencia en algunos estudios que se han realizado y en los que no tomaron en consideración estos aspectos.

Si el ajuste necesario para llevar a cabo una estimación válida de la calidad de la atención requiere de un número potencialmente elevado de variables, esto obliga a la utilización



de técnicas estadísticas de análisis multivariado, con la finalidad de intentar controlar una estructura muy compleja de índole multicausal.

Existe una corriente entre los investigadores en el campo de la evaluación de la calidad de la atención, especialmente provenientes del campo clínico, en relación con la conveniencia de que los estudios en esta área deberían enfocarse de manera primordial en el proceso de la atención y no en los resultados en salud obtenidos por los pacientes. Quienes defienden este enfoque argumentan la necesidad de considerar muchos otros factores, además de aquellos inherentes a la atención médica, cuando se trata de explicar y entender los resultados que se observan en los enfermos. Los factores de índole social y económica, además de aquellos inherentes a la biología individual del paciente, intervienen también de manera decisiva, a veces, en la producción de un resultado en salud. Por tal razón y ya que sólo los elementos del proceso de atención pueden ser controlados por el personal de salud, solo éstos, entonces, deberían considerarse al momento de evaluar la calidad de la atención. Este enfoque se basa fundamentalmente en una perspectiva clínica.

Existe una corriente de opinión opuesta que impulsa la realización de estudios para evaluar la calidad de la atención, enfocados en el resultado obtenido por el enfermo. Quienes se muestran a favor de esta posición, señalan como argumento la carencia de relación identificada en ocasiones entre proceso y resultado, así como la falta de datos que avalen la eficacia de las medidas utilizadas durante el proceso de atención. Esta tendencia, en general, es defendida desde la perspectiva de la epidemiología y tiene un enfoque social de manera predominante.

## Garantía de calidad

Si se considera el tratamiento en su atención más amplia, referida a cualquier maniobra (terapéutica o profiláctica) dirigida a mejorar una condición, se observa que las estrategias para mejorar la calidad de la atención incluyen un considerable espectro de actividades. Este conjunto de maniobras se ha denominado de manera genérica como **garantía** o **aseguramiento de la calidad**. Desde un punto de vista individual, se proporciona garantía de calidad cuando se asegura a un paciente que los servicios médicos que recibe cumplirán con ciertos estándares.

## Modelos de garantía de calidad

Se han descrito dos enfoques para proporcionar garantía de la atención médica. El primero de ellos enfatiza el mejoramiento de la estructura y el proceso de atención médica; este enfoque se utiliza preferencialmente en el Sistema de Salud de Canadá. El segundo enfoque prioriza el resultado de la atención médica y de manera principal se identifica con el sistema de salud de Estados Unidos.

A fin de vigilar el cumplimiento por parte de los servicios de salud con ciertos estándares de calidad a nivel hospitalario, se fundó la Comisión de Acreditación de Hospitales en Estados Unidos (JCAH) en 1951. El objetivo de esta Comisión ha sido ofrecer algún



tipo de acreditación a los hospitales, que cumplen con ciertas guías predeterminadas de calidad.

Los objetivos centrales de un programa de garantía de calidad son: enfocarse en el proceso de tomar decisiones de salud para mejorar los resultados obtenidos por los pacientes; aumentar la responsabilidad profesional y la capacidad para mejorar la atención; utilizar la práctica clínica como una fuente de información para mejorar la calidad; mostrar que se mejora la calidad al evitar los problemas de sobreuso o utilización indebida o subutilización de los servicios que se proporcionan durante la atención médica, y disminuir los problemas derivados de un pobre rendimiento técnico e interpersonal por parte del personal de salud.

Los modelos de calidad de la atención médica probablemente no sean útiles en todos los lugares ni en todas las situaciones donde se presta este tipo de atención. Por ejemplo, el modelo clásico de Donabedian de estructura, proceso y resultado, al aplicarse en determinadas situaciones, ha causado que la garantía de calidad a veces parezca punitiva y excesivamente regulatoria. Por otra parte, los defensores del modelo de proceso de atención y quienes argumentan a favor de un modelo basado en los resultados, han producido un debate que algunos juzgan improductivo, aun cuando parece claro que el consenso se inclina a pensar que para una adecuada evaluación de la calidad de la atención siempre debe incluirse en el resultado la información obtenida de los pacientes.

En general, se han propuesto dos estrategias para mejorar la calidad. Una de ellas se deriva de la denominada “teoría de las manzanas echadas a perder”; ésta se basa en la creencia de que la calidad se incrementa en la medida en que se identifican los malos elementos y se les remueve del conjunto; se ha denominado también **método para obtener calidad por inspección**. Algunos componentes de su metodología se basan en la recertificación del personal de salud, identificación de valores extremos y de procesos adversos tanto de mortalidad como de morbilidad. A partir de este enfoque se señala que los problemas de calidad se deben a elementos deficientes y, por tanto, ocurren por pobre rendimiento del personal que proporciona los cuidados de salud. Se afirma que la causa de los problemas son las personas, su incompetencia, escasa precaución, etc. A partir de este enfoque se propone premiar o castigar para controlar a los individuos participantes en el proceso.

La segunda estrategia se deriva de la denominada “teoría de la mejora continua” y su desarrollo ocurrió inicialmente en Japón; a ésta se atribuye gran parte del éxito económico observado en dicho país en los últimos decenios. Con base en este modelo se ha identificado que los problemas y, por ende, las oportunidades para mejorar la calidad, se desarrollan a partir del proceso de producción y que los efectos en la calidad rara vez pueden atribuirse a falta de destrezas o a falta de motivación entre la gente involucrada en el proceso.

En consecuencia, el problema no es de esfuerzo, sino de diseño deficiente del trabajo, falla en el liderazgo o un propósito poco claro. De acuerdo con esta teoría, el mejoramiento de la calidad depende de la comprensión y revisión del proceso por parte del personal involucrado y se fundamenta en los datos proporcionados por el proceso mismo.

Entre las estrategias para obtener garantía de calidad a nivel operacional se encuentran técnicas como las actividades educativas, así como mecanismos de regulación como el imponer castigos financieros o programas de exclusión de personal. Algunos métodos indirectos se basan en creencias acerca de la competencia y de las fuerzas del mercado en la atención a la salud. De cualquier forma, parece aceptarse que ningún programa de garantía de calidad puede ser exitoso sin una mezcla de enfoques. En la actualidad, la mayoría de las técnicas propuestas no ha proporcionado evidencia suficiente para mostrar una clara superioridad, ni se conoce con el detalle suficiente su papel en la corrección de problemas identificados en la calidad de la atención proporcionada.

## Perspectivas de la calidad de la atención

Diversos son los retos para los que se requiere la formulación de estrategias que permitan enfrentar y resolver los problemas en el campo de la calidad de la atención. A continuación se mencionan algunas de las más importantes.

### Incorporación de las preferencias de los pacientes

Es necesario considerar el hecho de que los valores y las preferencias del paciente no siempre se incorporan en el proceso de tomar decisiones clínicas. Para conseguir este objetivo es indispensable una estrategia nacional que permita incrementar la cultura médica en la población. Sólo de esta manera los enfermos podrán involucrarse con responsabilidad en las decisiones relacionadas con su propia atención y, en consecuencia, serán capaces de vigilar y exigir que los estándares mínimos de calidad de la atención les sean proporcionados.

En este sentido, debe reconocerse la legitimidad de que los pacientes difieran en sus preferencias, acerca de los tipos de servicios que se les proporcionan y de los resultados que en su salud, razonablemente, pueden esperar. Este aspecto de la práctica médica alcanza tal importancia y a la vez tal complejidad que ninguna de las preguntas que se deriven de su planteamiento podrá resolver de manera inmediata este problema. ¿En qué medida los médicos deben tomar en cuenta a los individuos al momento de tomar decisiones referentes a su cuidado de salud? Cuando los valores y las preferencias de los pacientes individuales entran en conflicto con los valores y las preferencias sociales, ¿qué debe tener precedencia?, ¿cómo compaginar una autonomía creciente por parte de los enfermos y que entra en conflicto con la beneficencia médica?

En este punto, puede resaltarse la unión indisoluble entre la calidad de la atención médica y la ética de la profesión. El médico encara la toma de decisiones ante los enfermos individuales y tiene que equilibrar sus obligaciones de beneficencia (el deber de hacer el bien a sus pacientes) y su obligación de no dañar, con los derechos de autonomía del paciente (el derecho del paciente a tomar sus propias decisiones en forma independiente); asimismo, ha de tomar en consideración aspectos de equidad y justicia (el deber de no discriminar entre grupos de personas y de distribuir los recursos juiciosamente, sin elementos de arbitrariedad ni gusto personal).

## Epidemiología de la calidad

Es necesario emprender el conocimiento de la epidemiología de la calidad en las instituciones de salud en México, incluyendo la práctica privada de la Medicina. Asimismo, es claro que la existencia de variaciones en los patrones de la práctica médica tiene repercusiones sobre los resultados que se obtienen en los pacientes. Determinar la magnitud y forma en que estos patrones varían, así como los factores que influyen sobre dicha variación, constituye un aspecto fundamental para la evaluación de la calidad de la atención prestada. Es conveniente mencionar que en México hay todavía escasos estudios, hasta la fecha, en estos aspectos.

Una crítica a los modelos de garantía de calidad es que son muy escasos los estudios controlados, que se han utilizado para evaluar la utilidad de dichas medidas. Por otra parte, estos programas generalmente no han tomado en cuenta la determinación del tamaño de la muestra suficiente para alcanzar conclusiones válidas.

## Promoción de la investigación de la calidad de atención

Es evidente que la promoción de investigación en esta área resulta indispensable para obtener información válida, que ayude a tomar decisiones para mejorar el sistema de salud. Se requiere, por otra parte, el desarrollo de métodos que permitan evaluar tanto la calidad de la atención que se otorga como el impacto de las medidas promovidas por los programas de garantía de calidad. Además, es necesario evaluar de manera rigurosa la utilidad de la aplicación de estas técnicas, incluyendo las propuestas por los programas de mejoramiento continuo de la calidad, especialmente en lo relativo a su impacto sobre los procesos de atención y no sólo en lo que se refiere a la estructura.

En este aspecto, no basta la demostración de su utilidad en otras áreas del mundo; parece indudable que las características culturales pueden influir de modo decisivo en la respuesta del personal de salud a dichos programas. Se requiere el desarrollo de métodos de difusión y evaluación de los programas, que se consideren necesarios para mejorar el rendimiento de los profesionales que atienden la salud. Desde luego, es indispensable la formación de personal entrenado en técnicas de evaluación, garantía de calidad e investigación, que permita la realización y evaluación de esta metodología. Algunos esfuerzos se han realizado en la pasada década en México, incluyendo una especialización en Calidad de la Atención Médica.

Donabedian ha sugerido la necesidad de que surja un nuevo tipo de profesional, el epidemiólogo en competencia clínica, quien debe tener destrezas epidemiológicas y en administración de la atención de la salud; además, el autor de este capítulo agregaría experiencia clínica. En diversos hospitales del mundo, fundamentalmente en Estados Unidos, han surgido programas para la evaluación de la calidad de la atención; dichos programas en general se han fijado tres aspectos por evaluar: excelencia en la atención proporcionada, evaluación de costos y satisfacción del paciente.

No ha de olvidarse, sin embargo, que tales programas requieren sustentarse en actividades académicas que deben formularse en el ámbito de las universidades. Estos pro-

gramas hospitalarios, por lo demás, tienen que ser suficientemente flexibles y formales, de manera que puedan adaptarse a diferentes condiciones y sistemas de salud.

## Desarrollo de instrumentos de medición

El personal de salud puede tener una idea clara de la necesidad de medir la calidad de lo que se efectúa; el problema, a decir de algunos, es que aún no sabe cómo hacerlo. Es claro que la sociedad debe realizar los esfuerzos necesarios para incrementar la investigación relacionada con el desarrollo de instrumentos que permitan aproximarse a una medición tan compleja y a la vez tan retardadora. Algunas recomendaciones incluyen aumentar la realización de experimentos clínicos para comprobar la eficacia de las prácticas médicas, enfatizando especialmente lo que se refiere a la evaluación de la tecnología diagnóstica (proceso).

Como segundo paso, es necesario llevar a cabo estudios cuidadosos de cohortes en enfermedades comunes, con el objeto de determinar si los procesos de atención se relacionan con resultados aceptables en el paciente. En este sentido, se ha reconocido la necesidad de obtener datos en forma detallada de aspectos clínicos, de laboratorio y tratamiento; asimismo, también deben obtenerse mediciones tanto funcionales como psicológicas del resultado obtenido por el enfermo. En tercer lugar, ha de considerarse la relación costo-efectividad de dichas prácticas y, en cuarto lugar, habrá de requerirse desarrollar y evaluar mecanismos que permitan medir y aplicar medidas de garantía de calidad.

A corto plazo, en este último aspecto, los programas de garantía tienen que enfocarse en los métodos para simplificar los procedimientos; en la evaluación de la calidad existen pocos criterios sólidos de proceso, para las enfermedades más comunes, y deberán evaluarse los padecimientos en que la experiencia sugiere que puede haber deficiencias en la atención que se les presta. Por otra parte, si un proceso no ha sido evaluado mediante experimentos clínicos, pero existe la creencia de la comunidad médica de que es efectivo, deberá asumirse en principio y con cautela que lo es; la investigación ha de realizarse buscando demostrar si dicho proceso que ya se sabe eficaz, realmente lo es en determinado lugar y si esto no fuera así, hay que encontrar las razones por las cuales no resulta efectivo.

Dada la alta complejidad de la práctica médica, las diferencias que se observan entre las unidades de atención médica se explican por diversos factores, uno de los cuales, de gran importancia, se deriva de las diferencias entre las características de los pacientes que acuden o ingresan a dichas unidades. Las diferencias en gravedad requieren el desarrollo de instrumentos que permitan realizar los ajustes que resulten convenientes.

## Conclusiones

Los modelos de evaluación de la calidad de la atención y del aseguramiento de la misma se han desarrollado en países distintos a México, fundamentalmente en Estados Unidos y Europa. Antes de incorporar dichos modelos a la realidad mexicana se deben evaluar

con rigor y juicio crítico las ventajas y desventajas de su aplicación y han de identificarse los métodos y la forma en que pueden incorporarse al contexto mexicano.

Por otra parte, la contribución del médico mexicano no debe ser aceptar en forma tácita y pasiva lo que ha sido bueno para otros; es indudable que se requiere creatividad y talento para desarrollar nuevos instrumentos y mecanismos que permitan mejorar la atención que el personal de salud brinda a las personas de México.

No ha de olvidarse que el objetivo final es que el personal de salud, las instituciones de salud y la sociedad en general se convenzan de las necesidades y ventajas de contar con sistemas de evaluación de la calidad de la atención, que permitan la práctica de una medicina con un nivel de excelencia, para los niveles de conocimiento actualmente existentes.

La evaluación de la calidad de la atención o la auditoría clínica constituyen términos de carácter científico, que definen una disciplina rigurosa, necesaria, para evaluar los servicios de salud en su conjunto y en consecuencia son útiles para pacientes, personal de salud, administradores y políticos en el sentido de que pueden permitir una práctica más racional de la Medicina.

---

## Bibliografía

- Asch SM, Kerr EA, Keesey J *et al.* Who is at greatest risk for receiving poor-quality health care? *N Engl J Med* 2006;354:1147-56.
- Berwick DM. What "patient-centered" should mean: confessions of an extremist. *Health Affairs* 28, no. 4 (2009): w555-w565 (published online 19 May 2009; 10.1377/hlthaff.28.4.w555).
- Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. *N Engl J Med* 1996;335:966-70.
- Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care* 2000;12:281-295.
- Brook RH, Chassin MR, Fink A *et al.* A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53-63.
- Campbell SM, Roland MO, Buetow S. Defining quality of care. *Soc Sci Med* 2000;51:1611-25.
- Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. *Br Med J* 2003;326:816-819.
- Department of Health, National Health System. Assuring the quality of medical practice. Implementing supporting doctors protecting patients. 2001. [www.doh.gov.uk/assuringquality](http://www.doh.gov.uk/assuringquality)
- Donabedian A. Explorations in quality assessment and monitoring Volume 1: The definition of quality and approaches to its assessment. Michigan: Health Administration Press, Ann Arbor, 1980.
- Institute of Medicine. To err is human: building a safer health system. 1999. [www.nap.edu/books/0309068371/html](http://www.nap.edu/books/0309068371/html)
- Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. 2001. [www.nap.edu/books/0309072808/html](http://www.nap.edu/books/0309072808/html)
- Jencks SF, Huff E, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. *JAMA* 2003;289:305-312.

- Kirk SA, Campbell SM, Kennell-Webb *et al.* Assessing the quality of care of multiple conditions in general practice: practical and methodological problems. *Qual Saf Health Care* 2003;12:421-427.
- McGlynn EA, Brook RH. Evaluating the quality of care. En: Andersen RM, Rice TH, Kominski GF. Changing the U.S. Health Care System. Second edition. San Francisco, California; Jossey-Bass, 2001;150-182.
- OECD Health Policy Studies. Improving value in Health Care. Measuring quality. OECD 2010.
- Pilote L, Tager IB. Outcomes research in the development and evaluation of practice guidelines. *BMC Health Services Research* 2002;2:7.
- Sally G, Donaldson LJ. Clinical governance and the drive for quality improvement in the new NHS in England. *BMJ* 1998;317:61-5.
- Wendler D, Emanuel EJ, Lie RK. The standard of care debate: can research in developing countries be both ethical and responsive to those countries' health needs? *Am J Public Health* 2004;94:923-928.
- Williams SC, Schmaltz SP, Morton DJ *et al.* Quality of care in U.S. Hospitals as reflected by standardized measures, 2002-2004. *N Engl J Med* 2005;353:255-64.



# Capítulo 17

## Selección del análisis estadístico

*Jorge Carreón García  
Laura Moreno Altamirano  
Guadalupe S. García de la Torre*

### Introducción

La participación de la Estadística ha sido muy importante en diversas ramas de la ciencia; en Epidemiología ha sido fundamental. La estrecha relación de la Estadística con el método científico hace de ella una disciplina imprescindible en la mayoría de los proyectos en el área de la salud.

Fue Pierre Charles-Alexandre Louis (1787-1872) el primer médico que utilizó métodos matemáticos para cuantificar variables presentes en los pacientes y sus enfermedades. En su libro *Méthode Numérique* publicó su clásico estudio sobre tuberculosis. En las propuestas de Louis para evaluar diferentes métodos de tratamiento están las bases de los ensayos clínicos que se hicieron un siglo después.

En Francia, Louis René Villermé (1782-1863) y en Inglaterra, William Farr (1807-1883), quienes habían estudiado Estadística Médica con Louis, fueron los primeros en hacer mapas epidemiológicos usando métodos cuantitativos y análisis epidemiológicos. Fue Francis Galton (1822-1911) quien, basado en el darwinismo social, fundó la Biometría Estadística.

Pero el cambio más fundamental en su aplicación a la Epidemiología se debe a Austin Bradford Hill (1897-1991), quien, con el ensayo clínico aleatorizado y en colaboración con Richard Doll (n. 1912), desarrolló el clásico estudio sobre la relación entre el tabaquismo y el cáncer de pulmón.

El pensamiento estadístico no sólo permite resolver las interrogantes metodológicas para dar respuesta a una hipótesis, sino que permite organizar la investigación desde el diseño general, diseño de muestreo, control de calidad de la información, análisis y presentación de resultados.

Este capítulo se concreta a proponer algunos lineamientos generales para el plan de manejo estadístico de la información producto de investigación.

Cuando un diseño de investigación es limitado y defectuoso no logra cumplir su propósito, ni siquiera si se cuenta con el análisis estadístico más refinado. De manera contraria, cuando el diseño cumple con los criterios de calidad y rigor metodológico, se obtendrán mejores aproximaciones al problema que se estudia. De esta manera, el método de análisis de la información adquiere importancia mayúscula: si los datos son la “materia prima”, y los resultados el “producto”, el análisis de la información constituye el “proceso”.

La generación de información en todas las áreas, incluida por supuesto la Medicina, en la actualidad es de tal magnitud que se ha hecho indispensable el conocimiento de ciertos procedimientos estadísticos especializados para su comprensión y comunicación. Este conocimiento con frecuencia no es del dominio del médico clínico, y su desconocimiento lo limita para comprender la bibliografía médica, revisar críticamente la literatura y para contestar preguntas de investigación. Es decir, sin la formación mínima sobre el proceso de análisis de datos, el médico no procesa todo lo que lee, y la comunicación de sus hallazgos es inadecuada o incompleta.

En la actualidad, con la ayuda de la computadora, el procesamiento de información numérica y su análisis estadístico se han simplificado mucho. Sin embargo, como es evidente, la selección de la mejor prueba y la interpretación de los resultados continúan siendo materia exclusiva del juicio que ejerce un ser humano. No es posible (ni deseable) que las computadoras sustituyan al cerebro humano en procesos de tan grande importancia; por ello, es necesario seleccionar e interpretar la prueba de hipótesis estadística más adecuada, según los datos que se han obtenido.\*

Con el propósito de facilitar la selección del procedimiento estadístico, en este capítulo se propone un esquema de decisión para elegir el procedimiento o los procedimientos que puedan aplicarse a los datos que se desea analizar. Para comprender las preguntas que se plantean ante la selección de la prueba de hipótesis, es necesario aclarar algunos conceptos.

## Estadística

La Estadística es una herramienta útil en la obtención, organización, análisis y presentación de datos; no constituye por sí sola un recurso que garantice que la información se ha obtenido de forma adecuada. Para hacer uso correcto de esta herramienta, es necesario que antes del análisis se tengan claros los objetivos y las hipótesis de investigación, ya que son los ejes fundamentales para la correcta selección de la prueba estadística.

La Estadística se divide en descriptiva y analítica. La primera es útil para expresar de manera sintética las medidas de resumen: tasas, razones, proporciones o tendencia central (media, mediana o moda) y de dispersión (desviación estándar, varianza, percentiles y rango). Para elegir la medida por usar, primero se deben clasificar las variables según la escala de medición (o punto de vista estadístico) y así seleccionar la medida o medidas

---

\* Prueba de hipótesis es la forma estadística de comparar dos o más poblaciones.



correspondientes. La estadística analítica se utiliza para buscar diferencias significativas o no, entre la(s) característica(s) evaluada(s) entre dos o más grupos.

Para seleccionar la mejor prueba por aplicar se parte de dos supuestos:\* 1) La muestra fue obtenida de una población con distribución normal\*\* (aunque la muestra no tenga distribución normal) y 2) se obtuvo mediante técnicas aleatorias. Cuando estos principios no se cumplen, es necesario usar pruebas de hipótesis no paramétricas. Las estadísticas paramétricas de prueba son más potentes; sin embargo, hay ocasiones en que se necesita utilizar una prueba menos potente que requiere mayor tamaño de muestra; en cualquier estudio es aconsejable obtener el mayor tamaño de muestra posible. En el cuadro 17-1 se muestra un listado de pruebas estadísticas.

## Clasificación de variables

En el capítulo 9 se estableció que la clasificación de variables puede realizarse de acuerdo con la escala de medición: cualitativas (nominales u ordinales) o cuantitativas (discretas o continuas). La mejor forma de medición es la escala cuantitativa; una variable medida en dicha escala puede modificarse y convertirse en ordinal o nominal, es decir, cualitativa. Por ejemplo, la glucemia es una variable medida en escala cuantitativa (mg/100 ml) y una serie de datos se puede resumir en promedio (medida de tendencia central) y desviación estándar (medida de dispersión). Además, para los fines específicos de un estudio, esta variable cuantitativa puede convertirse en cualitativa ordinal, tomando límites clínicos o estadísticos (cap. 7), y así se clasifica a los individuos en grupos de glucemia baja, intermedia y alta, o bien de forma cualitativa nominal, si se dividen en categorías de normoglucémicos e hiperglucémicos. Estas conversiones son válidas; sin embargo, las inversas no lo son.

Una variable cualitativa nominal —como el tipo de suturas quirúrgicas— no puede tener números asignados a cada categoría y después promediar los valores. Asimismo, una variable medida en escala cualitativa ordinal, como serían los grados de quemaduras (primer, segundo y tercer grados), no se puede sumar ni promediar; no es válido concluir que la muestra estudiada tenía en promedio quemaduras de 2.5 grados.

La escala de medición elegida debe ser la que mida más adecuadamente la variable; un ejemplo que ilustra lo mencionado es la composición corporal, cuya medición es posible efectuar por medio de apreciación visual, en donde sólo se obtendría una variable cualitativa ordinal (buena, regular, mala), o mediante la relación de peso y talla (valores obtenidos directamente del individuo), en cuyo caso se puede obtener información en escala cuantitativa continua. Otro ejemplo es la tensión arterial, que puede medirse con escala cuantitativa continua (si se estima en milímetros de mercurio), o si se clasifica a los individuos en normotensos o hipertensos, se trataría de una variable cualitativa

\* “Supuesto” es la base que, de ser cierta, indica que el modelo que se aplica es el adecuado.

\*\* Cuando en este texto se menciona el término “normal”, se hace referencia a la distribución de Gauss o normalidad estadística y no necesariamente a la normalidad clínica.

**Cuadro 17-1.** Lista de pruebas estadísticas

|                                                                                                                 |                                                                                                                                                               |
|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A) Binomial</b><br>Chi cuadrada                                                                              | <b>M) <math>r</math> de Spearman</b><br>D de Summer<br>Tau de Kendall<br>Tau b, c                                                                             |
| <b>B) Mc Neman</b><br>Chi de Mantel y Haenszel<br>Chi para nivel de confianza                                   | <b>N) <math>t</math> de Student</b><br>Prueba de Z                                                                                                            |
| <b>C) Prueba exacta de Fisher</b><br>Chi cuadrada<br>Riesgo relativo<br>Razón de momios                         | <b>O) Walsh</b><br>Aleatorización<br>$t$ de Student pareada<br>Prueba de Z pareada<br>Log Rank                                                                |
| <b>D) Q de Cochran</b>                                                                                          | <b>P) Aleatorización</b><br>$t$ de Student<br>Prueba de Z<br>Log Rank                                                                                         |
| <b>E) Chi cuadrada</b>                                                                                          | <b>Q) Anova</b><br>Turkey<br>Prueba de F                                                                                                                      |
| <b>F) Coeficiente de psi</b>                                                                                    | <b>R) <math>r</math> de Pearson</b><br>Correlación parcial                                                                                                    |
| <b>G) V de Cramer</b><br>Lambda<br>Coeficiente de contingencia<br>Q de Yule                                     | <b>S) Riesgos de Mantel y Haenszel</b><br>Riesgos pareados<br>Chi de Mantel y Haenszel<br>Análisis categórico<br>Análisis discriminante<br>Análisis factorial |
| <b>H) Kolmogorov-Smirnov</b><br>Prueba de las rachas                                                            | <b>T) Varianza de dos vías</b>                                                                                                                                |
| <b>I) Prueba de los signos</b><br>Wilcoxon                                                                      | <b>U) Covarianza</b><br>Análisis factorial                                                                                                                    |
| <b>J) Friedman</b><br>Varianza<br>Wilcoxon                                                                      | <b>V) Análisis factorial</b><br>Regresión múltiple<br>Correlación parcial<br>Durbin-Watson                                                                    |
| <b>K) Prueba de la mediana</b><br>Mann Whitney<br>Kolmogorov-Smirnov<br>Prueba de las rachas<br>Prueba de Moses |                                                                                                                                                               |
| <b>L) Extensión de la mediana</b><br>Kruscall Wallis                                                            |                                                                                                                                                               |

nominal, a partir de la cual no podrían asignarse números a las mediciones, ni realizar operaciones aritméticas, es decir, no se podría obtener promedio o moda, etcétera, de estos datos. La clasificación de variables según su escala de medición permite la conversión de datos “duros” a datos “blandos”, pero nunca en sentido inverso.

Otra forma de clasificar a las variables depende de la relación que tengan entre sí, y por ello a esta clasificación se le llama “por su relación” o desde el punto de vista metodo-

lógico. Así, a las variables se les clasifica como independientes o dependientes, también llamadas de causa o efecto, manejo o evolución, tratamiento o respuesta; en cualesquiera de estas denominaciones prevalece la idea de que la variable independiente es la que modifica o afecta a la dependiente. Si la tensión arterial (TA) es causa de una enfermedad vascular cerebral (EVC), entonces la TA es la variable independiente y la EVC es la variable dependiente. En otra situación, la masa corporal modifica la TA; en este caso, la TA es la variable dependiente y la masa corporal, la independiente. Aunque las variables se puedan clasificar de múltiples formas, las que consideran la escala de medición y su relación son útiles en particular para la selección del análisis estadístico y es posible clasificar una variable de ambas maneras simultáneamente. Para profundizar en el tema, consulte el capítulo 9. En el ejemplo anterior, la TA desde el punto de vista estadístico es una variable cuantitativa, mientras que desde el punto de vista metodológico es independiente (cuando se trató de la EVC) o dependiente (cuando se trató de la masa corporal).

## Potencia de la prueba

Cabe señalar que cuando se aplica la prueba de hipótesis con la escala de medición más adecuada, el resultado de la prueba es más potente. La potencia de una prueba es su capacidad para probar una hipótesis con un tamaño de muestra menor. Cuando se necesita mayor tamaño de muestra para probar una hipótesis, la prueba se considera menos potente. Es un error aplicar una prueba menos potente a datos obtenidos con escala de medición cuantitativa o dura.

## Objetivo del estudio

Cuando el objetivo del estudio es buscar diferencias entre grupos, ya sea porque pertenecen a distintas poblaciones o porque se han sometido a diferentes procedimientos diagnósticos o terapéuticos, lo que se pretende probar en sentido estadístico es la igualdad de los grupos, es decir, que “no existen diferencias y si las hay se deben al azar”. En caso de no comprobarse dicha igualdad, se acepta que los grupos son diferentes. Un ejemplo es la diferencia en la función ventilatoria en sujetos que viven al nivel del mar (donde hay mayor presión parcial de oxígeno) comparada con un grupo que vive en el altiplano. Otro ejemplo es la comparación de la respuesta de dos fármacos diferentes aplicados a dos grupos de pacientes, en el supuesto de que la respuesta a uno de los medicamentos será mejor.

En ocasiones el objetivo del estudio es comprobar la relación que hay entre dos variables como la ingestión calórica y el peso, o la respuesta a diferentes dosis de un fármaco. En algunos casos, la relación es directa, como la ingestión calórica y el peso (cuando aumenta la ingestión, es habitual encontrar aumento en el peso); en otros, la relación es inversa, como el uso de diferentes dosis de un medicamento beta bloqueador y la tensión arterial (al aumentar la dosis del fármaco disminuye la TA).

## Selección de individuos en estudio

Existen dos posibilidades de selección: 1) independiente, que se utiliza cuando la selección de cada sujeto no está influida por la selección de otro, y 2) dependiente, cuando se buscan, de manera intencionada, individuos con características específicas para realizar alguna comparación. Por ejemplo, cuando se desea comparar la respuesta a dos medicamentos “A” y “B”, por cada individuo de 30 años, de sexo masculino y 75 kg de peso, que reciba el fármaco “A”, se busca un individuo del mismo sexo y aproximadamente del mismo peso y edad; por tanto, la selección de los sujetos a quienes se administra el medicamento “B” no es aleatoria. Cuando se buscan relaciones, la selección de individuos es independiente y cuando se trata de hallar diferencias, la selección es dependiente.

## Número de variables

El tipo de análisis depende del número de variables por comparar. El análisis es simple cuando se relaciona una causa con un efecto, y es múltiple cuando se quiere establecer la asociación entre una causa y múltiples efectos, o múltiples causas y un solo efecto. El análisis múltiple se puede ejemplificar al estudiar la asociación entre hipertensión y obesidad; sin embargo, se sabe que esta relación está afectada por otras variables como tabaquismo, edad, sexo, actividad y concentraciones de colesterol, que también deben analizarse para la comprobación de las hipótesis.

Con base en los conceptos señalados, la forma de seleccionar la prueba o las pruebas más adecuadas implica responder las preguntas planteadas en el recuadro siguiente.

- ¿Qué tipo de análisis se desea efectuar? De acuerdo con el número de variables involucradas, el análisis es simple o múltiple.
- ¿En qué escala de medición está clasificada la variable dependiente? Cualitativa, nominal u ordinal o cuantitativa.
- Para el análisis múltiple, ¿en qué escala de medición están clasificadas las variables independientes y las dependientes?
- ¿La selección de los individuos se realizó de manera dependiente o independiente?
- ¿Cuántos son los grupos por comparar?

En la siguiente página se presenta un flujograma (figura 17-1) con las preguntas antes señaladas y al final de éste, una serie de opciones que deben buscarse en la lista de pruebas anexa. Por ejemplo, en un estudio donde se compara la eficacia analgésica de dos medicamentos (F y Z), la medición del dolor se hace en escala ordinal, usando la escala visual análoga; se clasifican en 10 grados de dolor y la selección de individuos es aleatoria para su asignación a cada grupo. Para comprobar que existe menor dolor con uno de los analgésicos en comparación con el otro, se selecciona la prueba con los parámetros mencionados en el recuadro de la página 282.

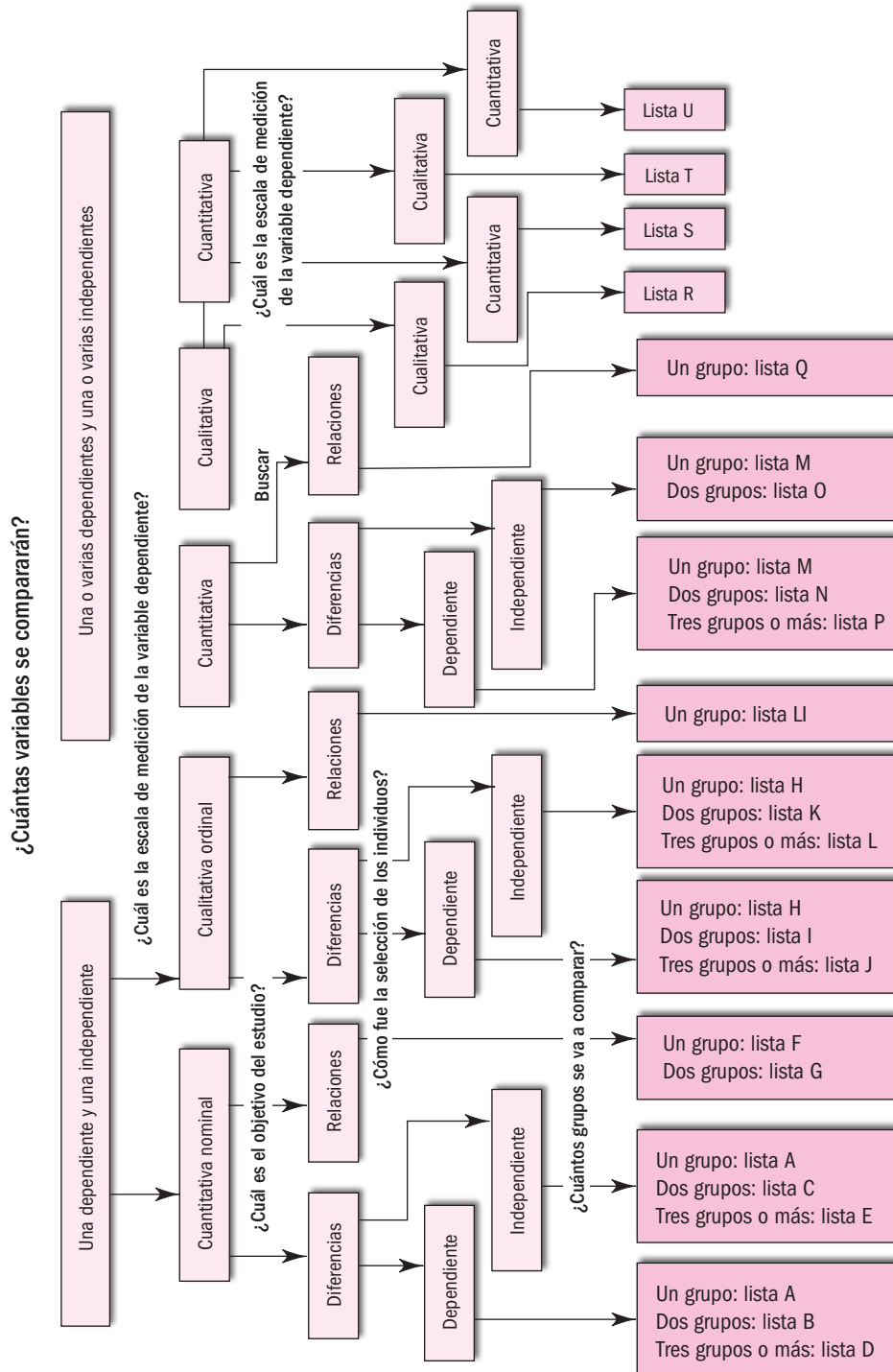


**Figura 17-1.** Selección de la prueba estadística.

- La comparación se efectúa con análisis simple, es decir, una variable dependiente y otra independiente.
- La escala de medición (de la variable dependiente) es ordinal.
- Se buscan diferencias entre los grupos.
- La selección de individuos es independiente.
- Son dos grupos.

Con estos elementos se revisa la lista de pruebas estadísticas (cuadro 17-1): de la mediana, de Mann Withney, de Kolmogorov, de las rachas y de Moses. En este mismo ejemplo, aunque el dolor se mide en escala ordinal, es factible clasificarlo en dos grupos: el de quienes presentan dolor de 0 a segundo grado y aquel de los sujetos que sufren dolor mayor de segundo grado; es decir, la medición del dolor se convirtió en una escala ordinal a una nominal y, en este caso, la selección de la prueba varió. El análisis sigue siendo simple; la escala de medición se cambia a nominal. En este caso, las pruebas sugeridas son Fisher, chi (ji) cuadrada y cálculo de riesgos. Para buscar los procedimientos específicos de cada prueba, se sugiere revisar la bibliografía.

---

## Bibliografía

- Kleinbaum DG, Kupper LL. *Applied regression analysis and other multivariable methods*. Duxbury Press, Boston, EUA, 1978.
- Levin J. *Fundamentos de Estadística en la investigación social*. Harla, México, 1979.
- Scheffler W. *Bioestadística*. Fondo Educativo Interamericano, México, 1981.
- Siegel S. *Nonparametric Statistics*. McGraw-Hill, Nueva York, EUA, 1956.



# Capítulo 18

## Significancia estadística y significancia clínica

*Juan José García García*

### Introducción

La toma de decisiones en la atención clínica respecto a la elección de una prueba para establecer el diagnóstico del paciente, prescribir una medida preventiva, terapéutica o de rehabilitación, proponer las acciones para modificar el pronóstico, entre otros aspectos de la atención al paciente supone, en cada caso, seleccionar la mejor opción.

Pero, ¿cómo se define que se trata justo de la mejor opción?

La respuesta, entre las diferentes facetas que pudiera tomar, enfrenta la noción de significancia.

### Significancia estadística

El análisis epidemiológico que se realiza, tanto en estudios observacionales como experimentales, exige la comparación entre un grupo de personas que presenta una característica respecto a otro que no la tiene.

En este contexto, las herramientas estadísticas que se utilizan para propósito del contraste, requieren de una secuencia de etapas que incluyen el planteamiento de un cierto criterio, definido antes de obtener los resultados, para establecer cuál será la probabilidad de que las diferencias observadas (o la asociación entre variables), si las hay, sean debidas al azar.

La selección del procedimiento estadístico a utilizar para analizar un problema determinado, depende de varios factores, entre los cuales se deben contar al menos los siguientes (figura 18-1):

- La naturaleza de los datos  
Tipo de variables. Cualitativas (nominales u ordinales) y/o cuantitativas (discretas o continuas).

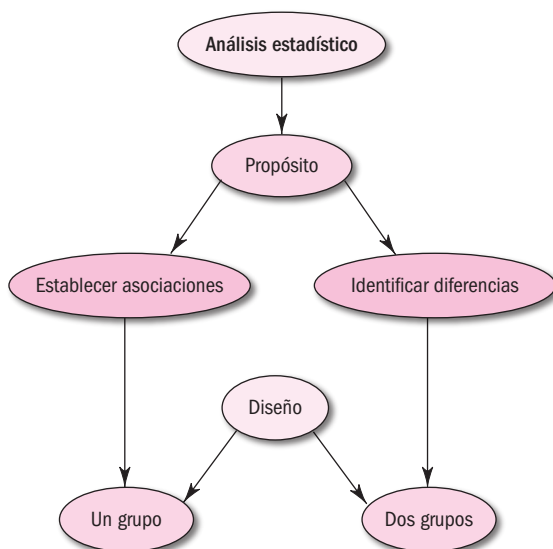


**Figura 18-1.** Propósitos del análisis estadístico relacionado con el diseño del estudio. Se asume que el hecho de abordar un solo grupo puede proporcionar únicamente medidas de resumen según el tipo de variables, en tanto que trabajar con dos grupos o más, incluye las posibilidades que ofrece tener uno solo y otras alternativas.

- El propósito que se persigue.  
Identificar asociaciones entre variables o diferencias entre grupos.  
El tipo de diseño del estudio.  
Número de grupos.  
Relación o independencia entre los mismos.

#### Etapas de la aplicación de una prueba estadística:

- Identificar la naturaleza de los datos.
- Plantear hipótesis estadísticas.
- Definir, en función de lo anterior, el tipo de prueba a realizar: unilateral o bilateral.
- Establecer el nivel de significancia.
- Según el caso, identificar el número de grados de libertad.
- Identificar el valor crítico para la prueba.
- Formular la regla de decisión para el rechazo de la hipótesis nula.
- Calcular el resultado de la prueba.
- Comparar el resultado con el valor crítico.
- Interpretar.



## Etapas de la aplicación de una prueba estadística

### Naturaleza de los datos

Implica identificar, por una parte, el tipo de variables —y, con ello, el nivel de medición alcanzado y las medidas de resumen empleadas—, y por la otra, el número de grupos que participaron, así como la independencia o no entre los mismos.

### Definir la prueba estadística a utilizar

La información misma que ha de analizarse define la prueba particular a utilizar; esto permite plantear qué tipo de técnicas pueden emplearse: paramétricas o no paramétricas.

Las primeras, que buscan hacer inferencias sobre valores poblacionales (parámetros), deben cumplir con ciertos supuestos o restricciones, como:

- Aleatoriedad en la selección muestral.
- Normalidad de la distribución de la variable.
- Igualdad de las varianzas entre los grupos.
- Independencia o no de las muestras.

Cuando las condiciones apropiadas no se cumplen, los resultados del análisis pueden ser válidos sólo para los datos muestrales, en cuyo caso las técnicas no paramétricas correspondientes constituyen el camino a seguir.

### Planteamiento de las hipótesis estadísticas

Las pruebas de significancia trabajan con dos hipótesis explícitas:

1. Nula o de no diferencias, preferentemente:  $H_0$ .
2. Alternativa (relacionada con la hipótesis de investigación):  $H_a$ .

La hipótesis nula:

- Es la hipótesis que debe “probarse”.
- Se establece con el propósito de ser rechazada.
- Puede decirse que es el complemento de la respuesta tentativa que el investigador ha dado a la pregunta problema del estudio.
- Al ser planteada debe contener una proposición de igualdad:  $=$ ,  $\geq$ ,  $\leq$ .

En el primer caso ( $=$ ), que es la forma más común de expresarse, por ejemplo, la comparación que se establece entre un valor muestral y un valor poblacional o esperado, o entre dos o más grupos, señala simplemente que éstos son iguales entre sí.

La forma en que aparece enunciada es, a manera de ejemplo:  $\mu_1 = \mu_2$ , si se comparan dos promedios;  $\pi_1 = \pi_2$ , si se trata de dos proporciones. En el caso particular del análisis sobre la posible participación de una variable como factor de riesgo (o protector), en que se compara un grupo expuesto con uno no expuesto, el valor nulo, es decir

de no asociación, lo constituye la unidad, al estar utilizando medidas como el riesgo relativo o la razón de momios, por lo que el enunciado de  $H_0$  puede ser:  $RR = 1$ .

La hipótesis alternativa:

- Es la proposición que se considera como cierta en caso de rechazar la hipótesis nula.
- Se identifica con la hipótesis de investigación.
- Se plantea en forma complementaria a la hipótesis nula, por lo que deberá ser tal que entre ambas se contemplen, de manera exhaustiva y excluyente, los valores posibles que los parámetros pueden asumir. En este sentido, indica que los valores comparados son diferentes entre sí, sin importar la dirección que tomen. La forma en que queda enunciada es:  $\mu_1 \neq \mu_2$ . Esto da lugar a que la región de rechazo de la hipótesis nula se encuentre en ambos extremos de una distribución teórica de frecuencias, denominándose entonces como una prueba bilateral. Tal es la forma habitual recomendada. En ciertas situaciones en que la dirección de las diferencias es anticipada y se plantea que un grupo tiene un valor mayor que otro, se puede elegir una prueba unilateral, en la que la región de rechazo de  $H_0$  se encontraría a la derecha o a la izquierda de la distribución.

Debe señalarse que las pruebas estadísticas en realidad sólo indican si la hipótesis es apoyada o no por los datos disponibles.

Cuando no se rechaza una hipótesis nula, no se concluye que ésta es verdadera.

La aceptación de una hipótesis no implica demostración.

## Nivel de significancia

Corresponde al valor asignado para alfa la probabilidad aceptada por el investigador de cometer un error tipo I. Habitualmente, de manera arbitraria, se establece un valor igual a 0.05. En una prueba bilateral, este 5% se encuentra distribuido en ambos extremos de la distribución, por lo que en cada uno se localiza el 2.5% del área bajo la curva. En ciertas circunstancias en las que la decisión a la que da lugar el resultado de la prueba puede tener serias repercusiones, por ejemplo, al definir toxicidad de un fármaco, el valor de alfa se llega a establecer en 0.01.

## Grados de libertad (gl)

Distintas distribuciones, como la  $t$  o la  $\chi^2$ , toman valores con respecto al tamaño de la muestra y, de hecho, hay una distribución para cada tamaño muestral.

En estos casos, el cálculo de los grados de libertad de la prueba constituye una etapa adicional para la localización del valor crítico a rebasar.

Para cada una de ellas la manera en que se calculan es diferente, por lo que sólo a manera de ejemplo se señala que en la comparación de dos promedios de muestras independientes con varianzas iguales, en donde se va a aplicar la prueba  $t$ , el número de grados de libertad es igual a la suma de los tamaños muestrales menos 2 ( $n_1 + n_2 - 2$ ).

A partir de ello, el valor crítico se localiza en el cruce del resultado obtenido (gl), con la columna del nivel de significancia elegido para una prueba unilateral o bilateral.

Cuando se trabaja con tablas para el cálculo de  $\chi^2$ , el número de grados de libertad se obtiene multiplicando el número de renglones (modalidades de la variable predictora) menos 1, por el número de columnas (modalidades de la variable de resultado) menos 1. El caso particular de las tablas de dos columnas y dos renglones ( $2 \times 2$ ), tiene 1 gl, de acuerdo con lo ya expresado. Se trata de una situación común en el análisis epidemiológico, en que la variable predictora constituye un supuesto factor de riesgo y se trata de una variable cualitativa nominal dicotómica que toma las modalidades expuesto o no expuesto, y la variable de resultado es la presencia o no de un daño a la salud.

### Identificar el valor crítico

El valor crítico constituye un punto de corte que permite establecer la región de aceptación y la de rechazo de la hipótesis nula. Dicho valor se localiza en función de los rubros anteriores. Unos cuantos de dichos puntos pueden ser preestablecidos conociendo las condiciones generales que los incisos previos marcan, pero muchos dependen del caso particular por analizar.

Así, por ejemplo, si se van a comparar los valores de dos promedios, esto supone que se está analizando una variable cuantitativa continua, y que ya se evaluó que se comporta como la curva normal (en función de los valores del sesgo y la curtosis). Bajo estas premisas se elige la prueba Z, si las muestras por comparar son grandes (digamos, mayores a 30 elementos). Considerando el caso habitual de una prueba bilateral, y un nivel de significancia de 0.05, automáticamente, el valor crítico para la aplicación de la prueba es 1.96, a ambos lados de la curva, ya que entre estos dos valores se encuentra el 95% del área central de la misma, y el 5% restante se encuentra distribuido en los extremos.

En el caso enunciado del análisis de tablas de  $2 \times 2$  en las que se emplea  $\chi^2$ , con un nivel de significancia de 0.05 y un grado de libertad, el valor crítico es 3.84. Esta prueba sólo toma valores positivos.

### Regla de decisión

Constituye un planteamiento en el que se comparará el valor calculado de la prueba con el valor crítico establecido para decidir si la hipótesis nula se rechaza o no.

Una forma de enunciar esta regla podría ser la siguiente:

Se rechaza  $H_0$ , si el valor calculado de la prueba Z es  $> 1.96$  o  $< -1.96$ , o

Se rechaza  $H_0$ , si el valor calculado de la prueba  $\chi^2$  es  $> 3.84$ .

Todos los valores posibles que la estadística de prueba puede asumir son puntos en el eje horizontal de la gráfica en la que, en función del valor crítico, se identifican dos zonas: de rechazo y de no rechazo de la hipótesis nula.

Los valores de la región de rechazo son aquellos que tienen menor probabilidad de ocurrir si la hipótesis nula es verdadera. Se debe, pues, rechazar ésta, si el valor calculado de la estadística de prueba es uno de los valores que se encuentran en dicha región.

Los valores de la región de no rechazo son aquellos que tienen una mayor probabilidad de presentarse. Se diría entonces que no hay elementos para rechazar la hipótesis nula si el valor calculado de la estadística de prueba es uno de aquellos que se encuentran en esta zona de la distribución.

La decisión de cuáles valores quedan en la región de aceptación y cuáles en la región de rechazo se establece por: el tipo de prueba (unilateral o bilateral), el nivel de significancia designado por la letra  $\alpha$ , que representa la probabilidad de rechazar una hipótesis nula verdadera, el número de grados de libertad cuando corresponda, y el valor crítico.

Un valor calculado de la estadística de prueba que cae en la región de rechazo se dice que es significativo.

Todos los pasos anteriores son necesarios para llegar a la expresión de que existen diferencias (o hay asociaciones) estadísticamente significativas, lo que indica que la probabilidad de que sean debidas al azar es menor al 5%.

Cuando se rechaza una hipótesis nula verdadera se comete un **error tipo I**.

Esto implicaría que, bajo una hipótesis nula de igualdad, se estaría afirmando de manera equivocada que dos grupos son diferentes entre sí con respecto a la medida comparada, cuando, estadísticamente, no es así.

Bajo el concepto anterior, por ejemplo, al cometer un error tipo I se diría que un factor (una variable) aumenta (o disminuye) el riesgo de que se presente una enfermedad, cuando no influye en algún sentido; o que un tratamiento es mejor que otro, sin serlo en realidad.

Cuando no se rechaza una hipótesis nula falsa se comete un **error tipo II**, cuya probabilidad de ocurrencia se designa por la letra  $\beta$ .

Incurrir en este tipo de error significa que al final se afirmaría que no existen diferencias estadísticas entre los grupos, cuando en realidad sí las hay.

Al cometer un error tipo II, entonces, se estaría afirmando, por ejemplo, que un factor (una variable) no influye incrementando o disminuyendo la probabilidad de que se presente una enfermedad, cuando en realidad sí lo hace; o que dos tratamientos son igualmente efectivos, sin serlo.

Durante el desarrollo de una investigación nunca se sabe si se ha cometido o no uno de estos errores (cuadro 18-1).

**Cuadro 18-1.** Ocurrencia de errores de tipo I y de tipo II

| Conclusión de la prueba           | Realidad                  |                            |
|-----------------------------------|---------------------------|----------------------------|
|                                   | Hipótesis nula cierta     | Hipótesis nula falsa       |
| Estadísticamente significativo    | Se comete un error tipo I | Conclusión correcta        |
| No estadísticamente significativo | Conclusión correcta       | Se comete un error tipo II |

## Decisión estadística

Este paso consiste en el rechazo o no rechazo de la hipótesis nula en función del punto en el que se encuentra el valor calculado de la estadística de prueba dentro de la distribución.

Cuando la hipótesis nula no es rechazada, tampoco se puede decir que se acepta, pues se puede haber cometido un error tipo II.

No obstante, el resultado de la estadística de prueba sólo es una parte de la evidencia que influye sobre la decisión clínica o administrativa.

Cuando la magnitud de las diferencias entre los grupos es grande, el número de individuos que se requiere para la conformación de la muestra es relativamente pequeño, en contraste con el caso de que dichas diferencias sean mínimas, pues esto exige tamaños muestrales mayores para dejar de manifiesto su existencia.

En cualquier caso, si la muestra o muestras han sido seleccionadas de manera aleatoria, la probabilidad de cometer ya sea un error tipo I o uno de tipo II, se reduce a medida que el número de individuos estudiado se incrementa. Esto da lugar a una disminución en el error estándar y a una mayor precisión de la estimación.

La información que ofrece la expresión “estadísticamente significativo” está limitada en esencia a señalar la probabilidad de que los resultados observados sean producto del azar. Por tanto, desde el punto de vista aplicativo, la construcción de intervalos de confianza, por una parte, y el cálculo de medidas de efecto ofrecen al médico más elementos para la comprensión de un fenómeno y la toma de decisiones (figura 18-2).

Cada vez cobra mayor importancia la evaluación no sólo de los aspectos relacionados con la efectividad de una medida de intervención, en términos de una medida de desenlace (como sobrevida, por ejemplo), sino de aquellos vinculados a la producción de efectos

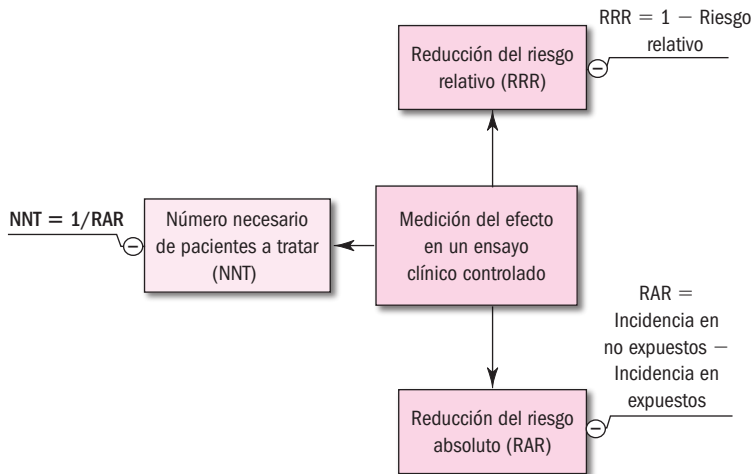


**Figura 18-2.** En los estudios de intervención pueden obtenerse diversas medidas que expresan el efecto logrado con una nueva maniobra, al compararse con un grupo de referencia. Se ejemplifican algunas de ellas y la forma de calcularse.

adversos, es decir, de la seguridad del producto, y de los costos, de adquisición, de administración, de atención a reacciones no deseadas, etcétera.

En el ámbito de la evaluación de medidas preventivas, terapéuticas o de rehabilitación, hablar de significancia clínica constituye una situación en la que los resultados observados a partir de un estudio apuntan o contribuyen hacia una transformación en las recomendaciones hasta ese momento aceptadas como la mejor opción, es decir, representan la base para proponer nuevos esquemas, nuevos patrones de prescripción, con las consecuentes ventajas para los pacientes.

Los cambios no ocurren de la noche a la mañana, pero si los datos son, digamos, espectaculares, llamarán la atención hacia la comunidad científica, y a la larga, la reproducción de resultados fortalecerá la evidencia de los beneficios ofrecidos.

La significancia clínica, entonces, va más allá del valor de  $p$ ; constituye una valoración más compleja en la que, además de la evidencia de los resultados reportados, entra en juego el juicio del médico, con base en su experiencia, ante una condición particular de un paciente.

### Ejemplo

Suponga que al evaluar el tiempo de sobrevida en pacientes con cáncer de colon metastásico se compara un grupo que recibió capecitabina con uno que fue manejado con Uracil/Tegafur (UFT), y que en el primero se alcanzó un promedio de 10 meses, mientras que en el segundo fue de 10.5 meses. Suponga que al aplicar la prueba respectiva de contraste se concluye que no hubo diferencias estadísticamente significativas en el tiempo de sobrevida entre ambos grupos, por lo que la probabilidad de que las diferencias observadas haya sido producida por el azar fue mayor a 5%. No obstante, cabría preguntarse, ¿qué representó para los pacientes de recibieron capecitabina esa discrepancia de 0.5 meses de sobrevida?

Transformada esa cifra en días, se estaría hablando de 15, es decir, dos semanas, tiempo durante el cual los pacientes, aun en etapa terminal, tuvieron ocasión, por ejemplo, de recibir la visita de amigos y familiares no vistos desde hacía mucho, reconciliarse con alguno de ellos, escuchar sus discos favoritos, leer un libro que habían dejado pendiente, recibir auxilios espirituales, etc. Para estas personas, haber formado parte de este grupo valió la pena, comparado con el que recibió UFT, aunque la estadística señale otra cosa.

## Bibliografía

- Cardiel MH. ¿Qué quiere decir el valor de  $p$ ? En: Calva MJJ (Director huésped). *Temas de Medicina Interna. Epidemiología clínica*. McGraw-Hill Interamericana, México, Vol II(2): 323-327. 1994.
- Cardiel RM. El clínico y la significancia estadística. En: Moreno AL, Cano VF, García RH. *Epidemiología clínica*, 2a. ed. McGraw-Hill Interamericana, México, 247-252. 1994.
- Colton T. *Estadística en medicina*. Salvat Editores, España, 103-158, 1979.

- Dawson B, Trapp RG. *Bioestadística médica*, 3a. ed. Editorial El Manual Moderno, México, 114-122. 2002.
- López-Jiménez F. Cómo medir la magnitud del beneficio de las intervenciones. Significancia estadística y clínica. En: López-Jiménez F, Obrador GT, Lamas GA. *Manual de medicina basada en la evidencia*. Editorial El Manual Moderno-JGH Editores, México, 163-177. 2001.
- Martínez-González MA. *Bioestadística amigable*, 2a. ed. Ediciones Díaz de Santos, España, 2006.
- Pagano M, Gauvreau K. *Fundamentos de bioestadística*, 2a. ed. International Thompson Editores, México, 2001.
- Wayne DW. *Bioestadística. Base para el análisis de las ciencias de la salud*, 4a. ed. Editorial Limusa, México, 2002.

# Capítulo 19

## Muestreo y cálculo de tamaño de muestra

*Juan José García García*



### Introducción

El término **muestreo** se refiere a la forma en que es seleccionada una fracción, llamada **muestra**, de la población que se desea estudiar.

En todo tipo de diseño epidemiológico debe contestarse, durante la planeación, una serie de preguntas a este respecto, es decir, a qué población se desea extrapolar los resultados, a quiénes se va a estudiar, bajo qué criterios serán elegidos los participantes, quiénes no son elegibles, cómo se llevará a cabo el proceso de selección, y cuántos deberán ser incorporados.

Estas preocupaciones tienen relación con la intención de obtener mediciones válidas (exactas) y confiables (precisas) acerca de los eventos de interés, lo cual supone que la manera apropiada de lograrlo es a través de procedimientos que permitan que todos los integrantes de la población de interés tengan, en general, una probabilidad igual y conocida de ser elegidos para ingresar al estudio.

La razón principal para el empleo de una muestra es el ahorro de recursos que implica la realización de un estudio, ya sea humanos, materiales o financieros.

#### **Ventajas del muestreo**

Bajo costo en relación con la enumeración completa de la población.

Menor consumo de tiempo.

Permite obtener mayor detalle de los datos.

Menor cantidad de personal requerido.

Mejor entrenamiento de quienes participan.

Puede ser la única opción para estudiar un problema.



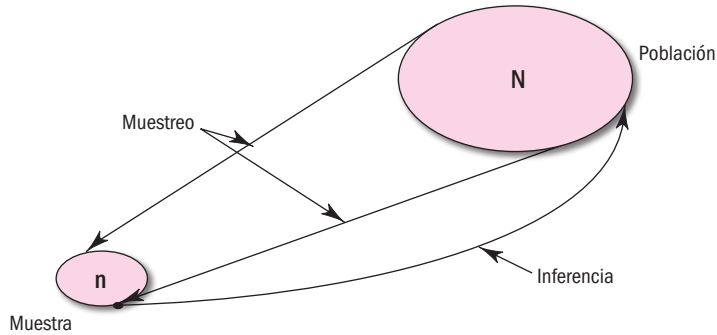


**Figura 19-1.** La población en su conjunto es representada por  $N$ , en tanto que la muestra se encuentra simbolizada por  $n$ . Las flechas que se dirigen de la población hacia la muestra representan el método de **muestreo** por el cual ésta ha sido seleccionada. Para propósitos de esta figura, se considera que dicho método ha sido de tipo probabilístico. Por último, la flecha que parte de la muestra hacia la población, se refiere a la tarea de estimar el comportamiento del fenómeno estudiado en la población en función de los hallazgos muestrales (**inferencia**).

## Población o universo

Se trata del conjunto a partir del cual se selecciona una muestra y sobre el cual, a su vez, se pretende hacer extensivos los resultados del estudio (figura 19-1). Este conjunto puede ser el de todos los habitantes de una región, todos los pacientes de un hospital, todos los pacientes con determinado padecimiento, todos aquellos que reciben un tipo de tratamiento, todos los que son sometidos a un procedimiento de diagnóstico. De igual forma, puede considerarse como tal a todos los expedientes del archivo clínico de la unidad médica, a los atendidos en cierto servicio, etcétera. Algunos autores hablan de **población blanco** para referirse a este conjunto, que suele ser de un tamaño tal que son insuficientes los recursos disponibles o el tiempo para estudiarlo.

### Población

Conjunto total de unidades (individuos, elementos, mediciones) existentes en un cierto lugar durante un periodo, que poseen una característica común que se desea estudiar. El número de elementos que la integran constituye el tamaño de la población. A menudo se le representa con la letra  $N$ .

El conocimiento más aproximado de los integrantes de la población implica disponer de un listado reciente de los mismos, lo que se conoce como **marco muestral**. Dicho marco contiene, entonces, una relación de las unidades de muestreo que, a su vez, son “colecciones no traslapadas de todos los elementos de la población”.

Un marco muestral, en mayor o menor medida, puede llegar a alejarse de la “verdadera población” en función de la actualización de información sobre la dinámica que

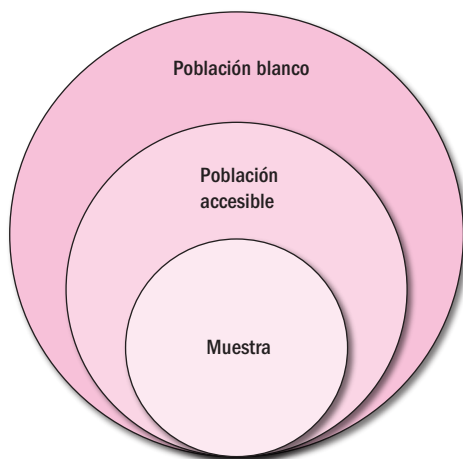


**Figura 19-2.** Debido a diferentes factores (espaciales, temporales, económicos, aceptación) del conjunto de la población blanco sobre la cual se desea tener alguna información, existe un subconjunto de ese universo que se encuentra accesible, a partir del cual se realiza la selección de los participantes de la muestra que efectivamente se estudia.

ésta tenga con respecto a ingresos y salidas, dados no sólo por nacimientos y defunciones, sino por fenómenos migratorios y, según el caso, por despidos o afiliaciones laborales, así como por referencia y contrarreferencia de pacientes.

Otro hecho importante es que, dada esa limitación en mantener la información al día, además de algunos aspectos adicionales relacionados con el acceso a los servicios de salud —ya sea públicos (asistenciales o de seguridad social) o privados—; el nivel de atención médica; la propia intensidad de la enfermedad, y la participación o influencia de la familia, esa población adquiere el adjetivo de **accesible**, para hacer referencia a aquella que ofrece ciertas facilidades para ser estudiada (figura 19-2).

Esto se hace aún más evidente cuando los individuos que se incorporan al estudio lo hacen de manera libre y consciente, dando su autorización por escrito.

## Muestra

Una muestra constituye una parte del conjunto llamado **población**. Se le representa con la letra **n**.

La conformación de este fragmento poblacional implica definir a partir del planteamiento del problema de investigación, y de la pregunta que se pretende contestar, qué característica(s) se requiere estudiar de los elementos (o unidades de observación seleccionados), tales características pueden ser variables cualitativas o cuantitativas, que serán representadas a través de las correspondientes medidas de resumen: razones, proporciones o tasas, en el primer caso, y medidas de tendencia central y de dispersión, en el segundo (figura 19-3). Estos aspectos dan lugar a la definición de los llamados *criterios de inclusión y de exclusión*.

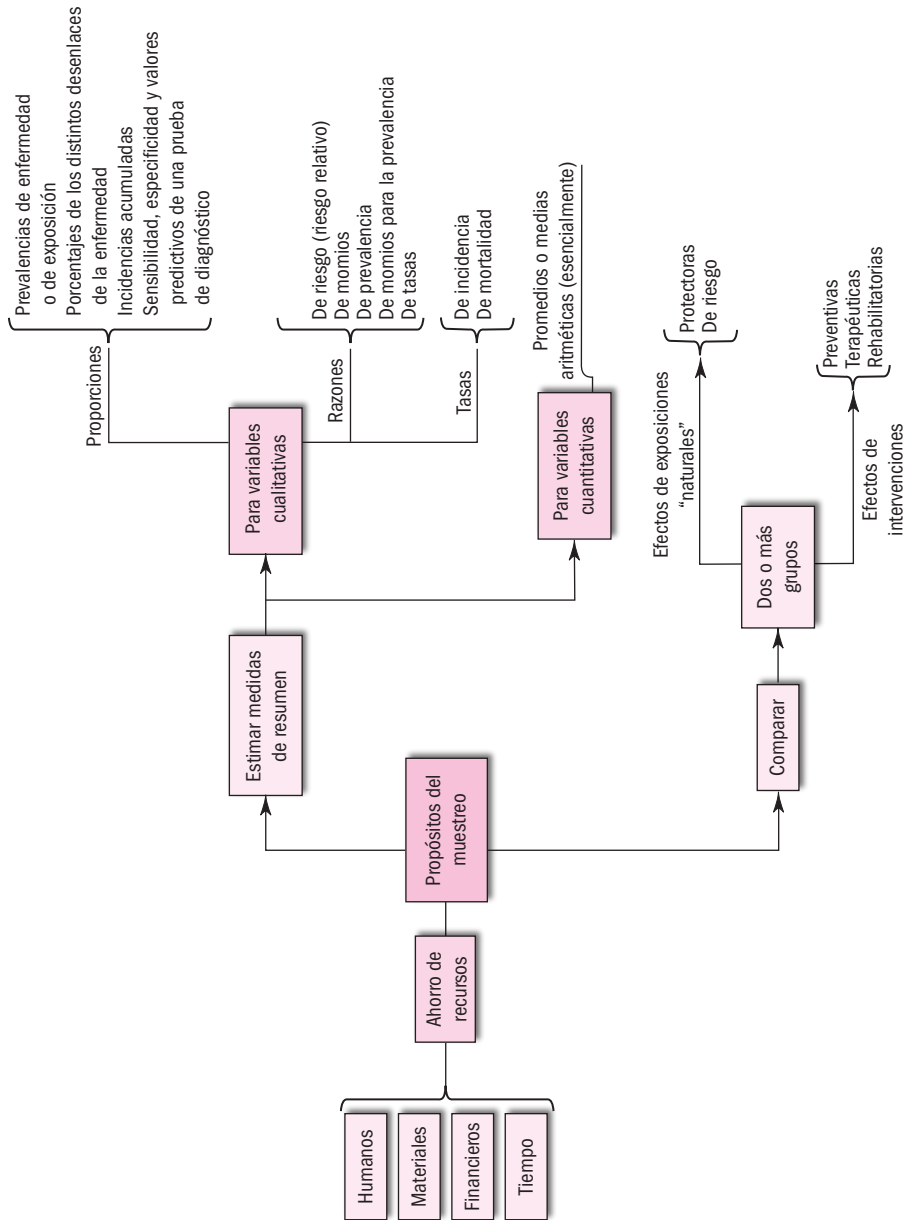


**Figura 19-3.** Los propósitos de obtener una muestra en los estudios epidemiológicos pueden agruparse en torno a la estimación de medidas de resumen según el tipo de variable, la comparación de grupos y el ahorro de recursos.

Hulley señala que los primeros “definen las principales características de la población estudiada que corresponden a la pregunta de investigación”, y agrega que éstas comprenden aspectos demográficos, clínicos, geográficos y temporales.

A manera de ejemplo, el interés puede estar en mostrar la distribución de los integrantes de la muestra según edad, sexo, gravedad de una enfermedad, o bien, algún valor antropométrico o bioquímico, en una región y tiempo dados.

Los criterios de exclusión, por otra parte, que han de ser tan pocos como sea posible, se refieren a subgrupos de personas que serían elegibles de no ser por características que pueden interferir con las mediciones o con el seguimiento, y por tanto, en la calidad de los datos obtenidos (véase capítulo 3).

Si se estudia una sola muestra, en general, se pretende, por medio de procedimientos estadísticos inferenciales, llegar a conclusiones sobre la población blanco, con base en principios probabilísticos. En este caso, se podrán construir **intervalos de confianza**, por lo general al 95% sobre los estimadores puntuales calculados en la muestra estudiada (figuras 19-1 y 19-2).

Un intervalo de confianza, genéricamente, se construye con el formato siguiente:  
 Estimador puntual  $\pm$  coeficiente de confiabilidad  $\times$  error estándar

Como ya se señaló, dada la naturaleza de las variables de interés, dichos estimadores serán de uno u otro tipo. A manera de ejemplos, en el ámbito clínico resulta importante tomar en cuenta aspectos como:

- a) Identificar la frecuencia con que se presentan y pueden esperarse las distintas manifestaciones clínicas de un padecimiento, la frecuencia de recuperaciones, complicaciones, secuelas, la probabilidad de sobrevivencia.
- b) El estudio de la capacidad de una nueva prueba de diagnóstico para reconocer correctamente tanto a los enfermos como a los sanos, y reducir al máximo el porcentaje de calificados en forma incorrecta, ya sea que se trate de enfermos no reconocidos por el examen realizado, o de sujetos sanos etiquetados como sospechosos de tener un problema de salud.

Si se estudian dos o más muestras, o subgrupos de una sola, es preciso realizar comparaciones en búsqueda de asociaciones entre variables o de diferencias que resulten estadísticamente significativas. Las técnicas utilizadas intentan medir la probabilidad de que dichas diferencias o asociaciones sean debidas al azar.

De manera simultánea, con un enfoque más epidemiológico que estrictamente estadístico, las pruebas de significancia se acompañan del cálculo de intervalos de confianza para los estimadores obtenidos.

- a) En este terreno se encuentran los estudios analíticos que buscan la existencia de una relación entre una característica y la ocurrencia de una enfermedad, de tal

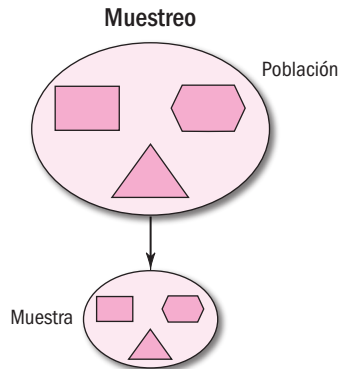


**Figura 19-4.** El empleo de métodos probabilísticos en la selección de la muestra pretende reducir la subjetividad del investigador en el proceso, así como lograr que la muestra “represente” de la mejor manera posible al conjunto del cual procede.

forma que pretenden establecer si se trata de un factor de riesgo o protector. Cada tipo de estudio epidemiológico ofrece distintas medidas de asociación.

- b) En el caso de la evaluación de medidas de intervención, sean preventivas, terapéuticas o de rehabilitación, el contraste entre dos o más grupos busca establecer cuál de las acciones, programas, medicamentos, tipos de cirugía, etcétera, ofrece mejores resultados.

Por otra parte, puede ser que el interés del investigador más bien radique en conocer las percepciones, experiencias, comportamientos, opiniones, valores, creencias, sentimientos (entre éstos se cuentan motivos de alegrías, angustias, miedos, enojos) o conocimientos de un paciente y/o su familia respecto a su situación particular de salud, el origen del problema. Así, el médico se interesará por aspectos como su situación de salud, el origen del problema, los riesgos de un procedimiento de diagnóstico y sus expectativas de recuperación, en cuyo caso, no es a través de indicadores, sino de empatía (“ponerse en el lugar del otro”) como se desea tener una aproximación al conocimiento y a la búsqueda conjunta de soluciones.

## Muestreo

Dado que, en general, la intención es extrapolar los resultados obtenidos en un segmento hacia el grupo del cual procede (población accesible) y, aún más, hacia un grupo más amplio que comparte características (población blanco), entonces cobra particular importancia el hecho de que la muestra sea lo más parecida posible a la población de origen (figura 19-4).

Lo anterior implica, esencialmente, eliminar o reducir el componente subjetivo que incorpora el investigador al tratar de definir quiénes serán los participantes de un estudio. Para ese efecto es para lo que se recurre a métodos a través de los cuales el azar se encarga de elegirlos (figura 19-5).

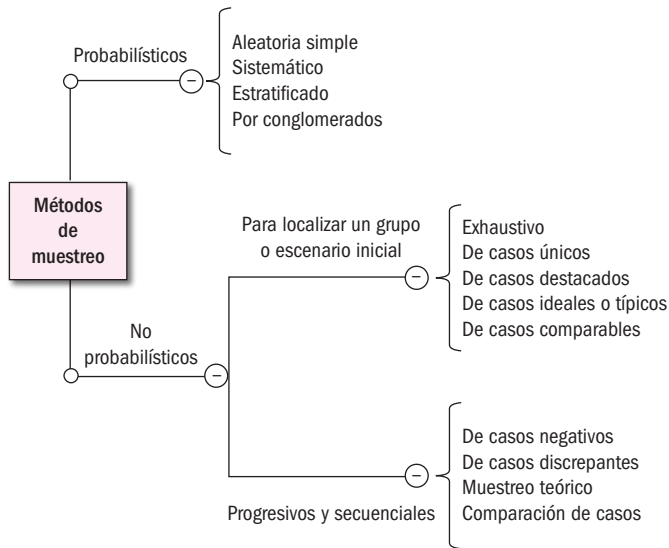


**Figura 19-5.** Resumen de los principales métodos de muestreo, ya sea probabilístico o no probabilístico. Estos últimos son particularmente empleados en investigación cualitativa.

## Muestreo probabilístico. Condiciones de aplicación

- Debe haber una probabilidad conocida de integrar a la muestra cada elemento de la población, sin que dicha probabilidad sea nula para alguno de ellos.
- Los procedimientos probabilísticos eliminan o, al menos, reducen la carga subjetiva del investigador en la selección de los elementos.
- La muestra debe exhibir internamente el mismo grado de diversidad que la población de la cual procede.
- El proceso inferencial siempre está sujeto a error. Una muestra puede considerarse representativa de ciertos aspectos de la población, dentro de un margen pequeño y conocido.
- El azar no necesariamente garantiza la representatividad, pero sí la imparcialidad.

### Muestreo probabilístico

Procedimiento estadístico para seleccionar la muestra a partir de la población, con el objeto de estudiar en ella alguna característica, y generalizar los resultados a la población de origen.

### Aleatorio simple

Cada elemento de la población tiene la misma probabilidad de ser incluido en la muestra; por lo anterior, se requiere numerar cada uno, de 1 hasta  $N$ .

## Muestreo sistemático

Supone, en primer lugar, numerar los elementos de la población entre 1 y  $N$ , y en seguida, calcular el espaciamiento de muestreo ( $k$ ), dividiendo  $N$  entre  $n$ .

El tercer paso es elegir al azar un número comprendido entre 1 y  $k$ , denominado **punto de arranque**, a partir del cual se obtendrán todos los demás, sumando el valor de  $k$  de manera consecutiva. Considere el siguiente ejemplo:

- Si  $N = 2\,000$  y  $n = 100$ , entonces el valor de  $k = \frac{2\,000}{100} = 20$
- Si el punto de arranque elegido al azar, entre 1 y  $k$ , es 15, el segundo número será 35, y así sucesivamente, se sumará de 20 en 20 hasta completar los 100 elementos que conformarán la muestra.

## Muestreo estratificado

Implica la identificación de características que dan lugar a la conformación de subgrupos homogéneos al interior pero heterogéneos entre sí. Los elementos dentro de cada estrato se seleccionan al azar. Los estratos de interés pueden identificarse en función de grupos de edad, sexo, nivel socioeconómico, intensidad de la enfermedad. El tamaño de cada subgrupo puede o no ser proporcional al tamaño del estrato de procedencia.

## Muestreo por conglomerados

Es útil cuando la población se encuentra dispersa. La selección de la muestra puede requerir varias etapas. En cada una de ellas se requiere del respectivo marco muestral, y la selección aleatoria de unidades muestrales.

En estudios de cobertura nacional, el muestreo polietápico puede implicar la selección de una entidad federativa, municipios, rancherías, colonias, barrios, unidades habitacionales, manzanas, viviendas, habitantes (figura 19-6).

En el caso tanto de estudios observacionales como experimentales (ensayos clínicos controlados), de tipo multicéntrico, esta modalidad de muestreo puede ser la opción que mejor responda a las necesidades del equipo de investigadores.

Las etapas pueden comprender la selección de países, regiones, entidades federativas, instituciones, hospitales, servicios, pacientes.

El esfuerzo que supone el muestreo probabilístico puede ser resumido en el grado en que se obtengan conclusiones correctas sobre lo que realmente ocurrió en el estudio (*validez interna*) y, por otro lado, en el grado en que tales conclusiones puedan ser aplicables a las personas (unidades de observación) no incluidas en el estudio (*validez externa*).

## Muestreo no probabilístico

En la investigación clínico-epidemiológica se presenta una alternativa importante para la selección muestral, en la que con un cuidadoso juicio subjetivo se puede alcanzar la

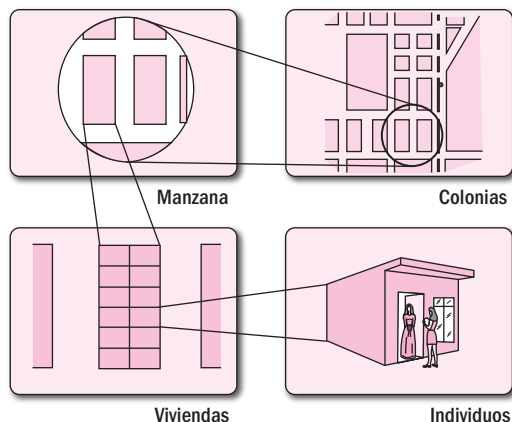


**Figura 19-6.** El muestreo polietápico implica la selección aleatoria de diversos niveles de organización de los espacios muestrales. Por ejemplo, revisando la figura de derecha a izquierda y de arriba hacia abajo, de una colonia se seleccionan manzanas, de éstas las viviendas y, en ellas, a los individuos que son de interés para el estudio. Méndez RI *et al.* 1987.

intención antes enunciada, a través de métodos en los que los individuos no son elegidos en forma aleatoria. El caso más claro al respecto lo constituye el llamado **muestreo por conveniencia**, en el cual la accesibilidad a ciertos sujetos define su participación. Este proceso puede incluir la incorporación consecutiva de aquellos que reúnen los criterios de inclusión, hasta alcanzar una cuota preestablecida.

Por otra parte, cuando el interés del investigador sobre un determinado problema no radica en la extensión de la información a obtener, ni en la generalización de las conclusiones que se desprendan, sino más bien en la profundidad de los datos, el estudio puede encontrarse en el terreno de la investigación cualitativa, y en este sentido se le presenta una gran diversidad de métodos para la conformación de la muestra (figura 19-5).

Rodríguez y colaboradores señalan dos grandes grupos de métodos de muestreo para este efecto:

- a) Aquellos dirigidos a localizar un grupo o escenario inicial, y
- b) Los que se consideran progresivos y secuenciales.

## Tamaño de la muestra

La definición del número de elementos a estudiar en un enfoque cuantitativo, estrictamente debe estar basada en lo siguiente:

1. **La variabilidad de la característica o fenómeno de interés en el universo que se pretende estudiar.** A mayor homogeneidad de la población, el tamaño muestral requerido será menor y, por el contrario, a mayor heterogeneidad, será necesaria una muestra más grande.



En el caso del estudio de una proporción, representada por la letra  $p$ , la variabilidad está dada por el producto de dicha  $p$ , por su complemento, llamado  $q$ , es decir, la proporción de individuos que no presentan la característica  $(1 - p)$ . Por ejemplo, si esa proporción a estudiar es la prevalencia,  $p$ , el valor de  $q$  está dado por la proporción de no enfermos.

Cuando la estimación que se desea se refiere a una media o promedio aritmético, la variabilidad corresponde al valor de la varianza.

El conocimiento de este dato, que será utilizado en las sustituciones correspondientes en la fórmula de cálculo por aplicar, ya sea manualmente o en un programa estadístico/epidemiológico, proviene de la revisión de la literatura sobre estudios similares al que se pretende realizar.

Si los grupos estudiados previamente son muy diferentes a los que se pretende abordar o que no exista información al respecto, hay dos soluciones prácticas:

Si se pretende estimar una proporción, se sugiere considerar que ésta es igual a 0.5, generando la máxima variabilidad posible, al multiplicarse por el complemento (0.5), alcanzando un tamaño muestral suficiente para conocer el valor de interés.

Si se pretende estimar un promedio, la recomendación es utilizar la información del estudio piloto que ha de realizarse de manera inicial para tener una aproximación al valor de la varianza.

2. **El margen de certeza que se desea obtener en la estimación.** El grado de confiabilidad habitualmente empleado es de 95%. Note que esto se refiere a la probabilidad con la que el valor estimado se encuentra dentro de cierto rango de valores. Cabe, pues, la probabilidad de que en 5% de los casos, el verdadero valor poblacional se encuentre fuera de esos límites.

En este sentido suele ser una constante pues, si, con base en el modelo de la curva normal, se utiliza el coeficiente  $Z$ , este número corresponde a 1.96. Si se desea una mayor confiabilidad, digamos de 99%, el valor de  $Z$  será de 2.58, que se reflejará en la mayor amplitud, también, del intervalo construido.

3. **La precisión que se desea alcanzar en los resultados.** En general, se espera que los datos se aglutinen cerca de la cifra encontrada en la estimación puntual del valor muestral de una proporción, un promedio, una medida de asociación. Esto habla de la distancia de un rango de valores alrededor de dicho estimador. A mayor precisión deseada, el tamaño muestral ha de ser mayor. Aunque el valor de la precisión puede ser establecido *a priori* por el investigador, por ejemplo, de 5%, su magnitud estimada a partir de los resultados es producto del coeficiente de confiabilidad a emplear por el error estándar o error de muestreo. Como es evidente, el valor de la precisión es la cantidad que se suma y se resta a un estimador puntual al construir un intervalo de confianza.

$$\text{Estimador puntual} \pm \text{coeficiente de confiabilidad} \times \text{error estándar}$$

El error estándar, o de muestreo, expresa el error aleatorio que se comete durante la selección de los participantes. Se calcula de diferente manera según el estimador que se

busca, o sea, una proporción, un promedio, una medida de asociación, y en su magnitud influye de manera sustancial el propio tamaño de muestra estudiado.

Así, entre más grande es la muestra, el error estándar es más pequeño, y el producto de éste por el coeficiente de confiabilidad da lugar a un número reducido, que al sumar y restar al valor del estimador puntual analizado, se traduce en una alta precisión, y viceversa.

### Tamaño de la muestra

Aspectos centrales a considerar:

- Variabilidad del fenómeno a estudiar
- Confiabilidad
- Precisión

Otros factores:

- Tipo de diseño de estudio
- Valores de alfa ( $\alpha$ ) y de beta ( $\beta$ )
- Magnitud de la diferencia entre los grupos
- Recursos disponibles
- Método de muestreo empleado

En el caso de la comparación de grupos, además de los aspectos antes señalados, cobran importancia dos valores más: el de la probabilidad de que en el proceso de investigación se pueda cometer un error tipo I, rechazar una hipótesis nula verdadera, y el de la probabilidad de cometer un error tipo II, el no rechazo de una hipótesis nula falsa.

El primero de estos valores es representado por la letra alfa ( $\alpha$ ), en tanto que el segundo, lo es por la letra beta ( $\beta$ ).

El valor de  $\alpha$  está asociado con el **nivel de significancia** empleado en las pruebas de contrastación de hipótesis y por lo general se establece con 5%, que representa la probabilidad aceptada por el investigador de cometer un error tipo I. Si éste ocurre (cosa que no va a saber el investigador), se traduce, dependiendo del propósito del estudio, en afirmar erróneamente que dos proporciones o dos promedios, provenientes cada uno de una muestra, son diferentes entre sí, cuando en realidad no lo son. Para fines prácticos podría estar señalándose de manera equivocada que la prevalencia de una enfermedad entre dos poblaciones es distinta; que la reducción en los niveles de colesterol en dos grupos que han recibido tratamientos distintos, también es diferente; que una característica o factor está asociado con la ocurrencia de una enfermedad sin que así sea, etcétera. El resultado erróneo reportado señalará que la probabilidad de que las diferencias observadas se deba al azar es menor al 5% ( $p < 0.05$ ).

El valor de  $\beta$  suele establecerse entre 10 y 20%. El hecho de cometer un error tipo II se traduce en situaciones tales como afirmar de manera equivocada que un factor no está

**Cuadro 19-1.** Relación entre realidad y resultado de la prueba

|                                  |                                                             | Realidad (población)                      |                                                  |
|----------------------------------|-------------------------------------------------------------|-------------------------------------------|--------------------------------------------------|
|                                  |                                                             | Existe diferencia o asociación (Ho falsa) | No existe diferencia o asociación (Ho verdadera) |
| Resultado de la prueba (muestra) | Diferencia o asociación significativa (rechazo de Ho)       | No hay error                              | Error tipo I                                     |
|                                  | Diferencia o asociación no significativa (no rechazo de Ho) | Error tipo II                             | No hay error                                     |

asociado con la ocurrencia de una enfermedad, cuando en verdad sí lo está; o concluir que no hay diferencias entre los resultados obtenidos por dos tratamientos, cuando en realidad sí los hay.

Cabe resaltar, por una parte, que el investigador al término del estudio no sabe si ha cometido alguno de estos tipos de error y, por otra, que la probabilidad de incurrir en ellos se reduce al aumentar el tamaño de la muestra (cuadro 19-1).

### Tamaño de muestra insuficiente

En el contexto del muestreo probabilístico, una situación que puede ocurrir es que el número de sujetos incluidos en el estudio resulte insuficiente para fines de generalizar los resultados y/o establecer la existencia de diferencias entre grupos o la asociación entre variables (exposición-efecto).

Si se trata de un solo grupo y se pretendía la estimación de un parámetro, el efecto de este hecho será la obtención de un intervalo de confianza amplio, es decir, poco preciso, que es resultado de un error estándar grande, a su vez producto de un tamaño muestral pequeño. *A posteriori*, si se deseara aumentar la precisión, tendría que ser a costa de sacrificar la confiabilidad.

En ocasiones, quizá no tan raras, el tamaño ideal de la muestra requerida no se logra satisfacer por distintas razones, entre las que destacan la limitación de recursos humanos, materiales, financieros o de tiempo disponible para tener resultados exigidos.

Otra razón sería partir de una estimación errónea de la frecuencia con la que se encontraría el problema de interés, y no se reúne una cantidad adecuada de casos.

En los estudios que implican seguimiento, un problema que se enfrenta es la pérdida de individuos por diferentes motivos, que puede comprometer seriamente la validez de los resultados si es mayor a 20%.

Además de las características del diseño del estudio, debe tomarse en cuenta el hecho de que por una u otra razón, puede haber una proporción de sujetos que no sean localizados o que no acepten participar, por lo que, para no perder las condiciones preestablecidas relativas a la confiabilidad y precisión, siempre es recomendable sumar al tamaño

de muestra calculado la proporción de no respuesta esperada. Por ejemplo, si se estimó una muestra de 100, y se supone que 15% no aceptará participar, éste se agrega al tamaño calculado, para que, en total, se estudien 115 personas.

Por otra parte, si se trata de un estudio comparativo, el efecto de un tamaño insuficiente de muestra se verá reflejado, por ejemplo, en el hecho de no poder concluir si un factor se encuentra asociado o no a un mayor o menor riesgo de enfermar, o bien, a no poder señalar si existe diferencia en la efectividad de dos medidas de intervención.

Si la magnitud de las diferencias entre los grupos es pequeña, se requiere un tamaño muestral grande para poder identificarlas. Lo contrario también es cierto, es decir, si las diferencias entre los grupos son marcadas, podrán reconocerse con tamaños muestrales menores.

Si el número de participantes es menor al previsto, pero ha sido seleccionado de manera probabilística, resulta de interés evaluar el *poder del estudio*, o sea, la probabilidad de detectar diferencias estadísticamente significativas entre grupos comparados. Esto se mide a través de la relación:

$$\text{Poder} = 1 - \text{beta}$$

De esto resulta que el poder de un estudio en razón aceptable es superior al 80%.

## Consideraciones finales

No existe un número mágico para un tamaño de muestra adecuado. En el caso de recurrir al muestreo probabilístico, ya se han señalado distintos factores que definen el valor calculado. Estos aspectos, inmersos en un tipo particular de diseño de un estudio, requieren fórmulas específicas a emplear. Así, a manera de ejemplo, las hay para la estimación de una proporción o de un promedio, cuando se trabaja con una sola muestra, y también las hay para la identificación de diferencias de proporciones o de promedios, si se trata de un estudio comparativo. Para este propósito, considere lo siguiente:

- a) Se pueden aplicar directamente, con menor o mayor grado de complejidad, realizando las sustituciones respectivas en los valores solicitados (cuadro 19-2).
- b) Es factible recurrir a tablas ya elaboradas de acuerdo con el caso. Distintos textos presentan ejemplos al respecto de diferentes situaciones. Los libros de Argimón, Schlesselman y de Fleiss incluyen amplios anexos relacionados con tamaño de muestra.
- c) Es posible emplear programas estadístico-epidemiológicos que contienen rutinas específicas. Por ejemplo, Epidat 3.1, un programa de distribución libre, desarrollado en Galicia, y apoyado por la Organización Panamericana de la Salud, incluye una opción para el cálculo del tamaño muestral, la selección aleatoria de los participantes y la asignación de sujetos a grupos de tratamiento, en el caso de estudios experimentales. Para el cálculo muestral ofrece dos grandes categorías, una en relación con la construcción de intervalos de confianza para diversos estimadores, y otra para el contraste de hipótesis, con diferentes tipos de diseños de estudio y

**Cuadro 19-2.** Fórmulas para el cálculo de un tamaño de muestra

Ejemplos:

$$\text{Estimación de una proporción} \quad n = \frac{pqZ^2}{d^2}$$

$$\text{Estimación de una media o promedio aritmético} \quad n = \frac{s^2 Z^2}{d^2}$$

$$\text{Comparación de dos proporciones} \quad n = \frac{\{Z_{\alpha} \sqrt{2pq} + Z_{\beta} \sqrt{[p_1q_1 + p_2q_2]}\}^2}{(p_1 - p_2)}$$

$$\text{Comparación de dos medias} \quad n = \frac{2(Z_{\alpha} + Z_{\beta})s^2}{d^2}$$

variados propósitos. El programa se complementa con extensa información de ayudas para cada uno de los métodos que ofrece (figuras 19-7 a 19-9).

Finalmente, a pesar de lo sencillo que puede ser trabajar con estas herramientas y hacer los cálculos correspondientes, en la medida que el proceso de investigación es un trabajo de equipo, siempre será recomendable que como parte del mismo se encuentre un integrante con la formación estadística necesaria para contemplar los diversos aspectos relacionados con la selección de los métodos más apropiados al caso, y la selección de los participantes.

En el caso de estudios comparativos, el tamaño muestral obtenido es para cada grupo.

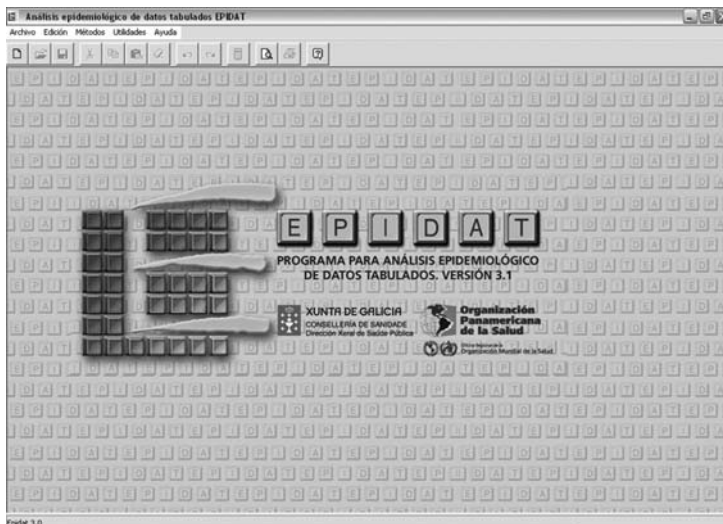


**Figura 19-7.** Pantalla principal del programa Epidat. Versión 3.1 (ya se encuentra disponible la versión 4.0).

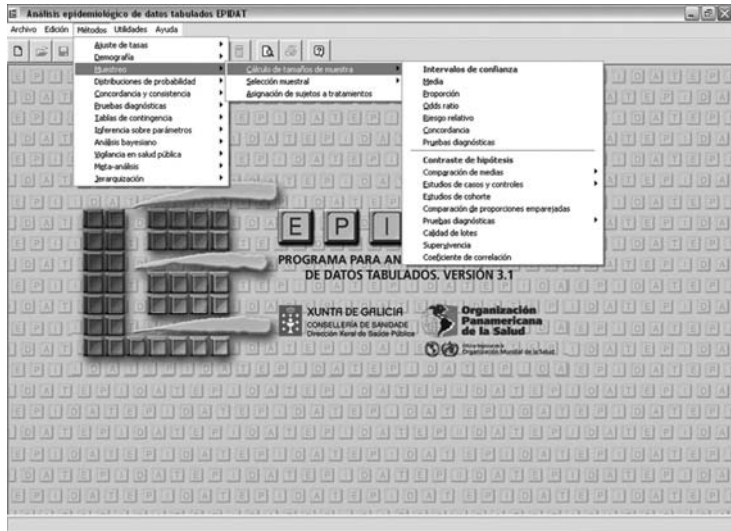


**Figura 19-8.** Se observan las ventanas relacionadas con las rutinas que ofrece el programa Epidat sobre cálculo del tamaño muestral. Otras opciones, no sombreadas, permiten la selección muestral a partir de una base de datos capturada en Excel, y la asignación aleatoria de sujetos a estudios de intervención.

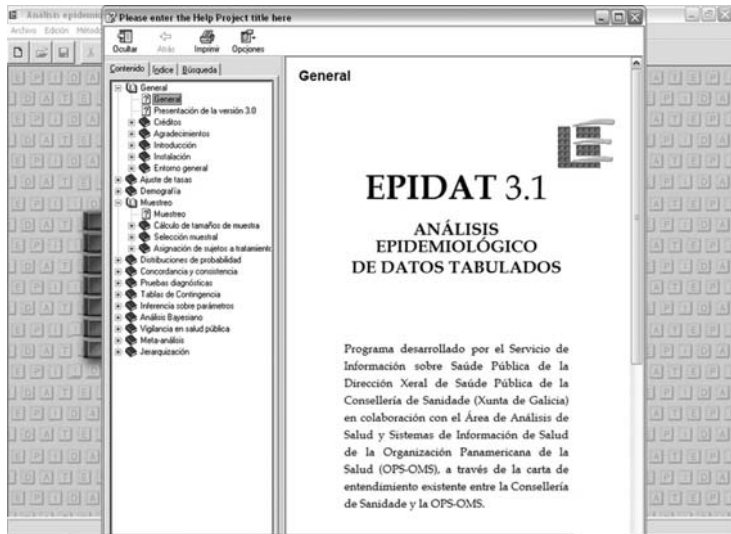


**Figura 19-9.** El menú de ayuda de Epidat presenta aspectos teóricos sobre el tema, así como la ilustración de la aplicación de estas técnicas a partir de archivos incluidos para tal propósito.

## Bibliografía

- Álvarez-Gayou J. *Cómo hacer investigación cualitativa. Fundamentos y metodología*. Editorial Paidós, México, 2003.
- Ardila MJM, Rodríguez MMN, Gil LJFA. Población y muestreo. En: *Epidemiología clínica. Investigación clínica aplicada*. Editorial Médica Panamericana, Colombia, 129-139. 2004.
- Argimon PJM, Jiménez VJ. *Métodos de investigación clínica y epidemiológica*, 2a. ed. Harcourt, España, 105-118, 2000.
- Argimón PJM, Jiménez VJ. *Métodos de investigación clínica y epidemiológica*, 2a. ed. Harcourt, España, 119-126, 2000.
- Browner WS, Newman TB, Hulley SB. Preparación para el cálculo del tamaño de la muestra: hipótesis y principios subyacentes. En: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Diseño de investigación clínica*, 3a. ed. Lippincott Williams & Wilkins, EUA, 57-96, 2008.
- Dawson B, Trapp RG. *Bioestadística médica*, 3a. ed. El Manual Moderno, México, 75-80, 2002.
- Dennis VRJ, Pérez MA. Cálculo del tamaño de muestra. En: *Epidemiología clínica. Investigación clínica aplicada*. Editorial Médica Panamericana, Colombia, 141-162. 2004.
- Fleiss J.L. *Statistical methods for rates and proportions*. John Wiley & Sons, Nueva York, EUA, 1981.
- Hernández SR, Fernández-Collado C, Baptista LP. *Metodología de la investigación*, 4a. ed. McGraw Hill, México, 235-270. 2008.
- Hulley SB, Newman TB, Cummings SR. Elección de los participantes en el estudio: especificación, muestreo y reclutamiento. En: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Diseño de investigación clínica*, 3a ed. Lippincott Williams & Wilkins, EUA, 31-42. 2008.
- Marrugat J, Vila J, Pavesi J, Sanz F. Estimación del tamaño de muestra en la investigación clínica y epidemiológica. *Med Clin (Barc)* 1998;111:267-276.
- Martínez-González MA, Bes-Rastrollo AAM. Estimación del tamaño muestral. En: Martínez-González MA (ed.). *Bioestadística amigable*, 2a ed. Ediciones Díaz de Santos, España, 373-417. 2006.
- Méndez RI, Namihira GD, Moreno AL. El protocolo de investigación: lineamientos para su elaboración y análisis. México: Trillas. 1987.
- Orozco MJA. ¿Cuánto es un número suficiente de individuos en una investigación? En: Calva MJJ (Director huésped). *Temas de Medicina Interna. Epidemiología Clínica*. Vol. II (2):337-349. 1994.
- Pagano M, Gauvreau K. *Fundamentos de bioestadística*, 2a. ed. Thompson Learning, México, 514-525. 2001.
- Rodríguez GG, Gil FJ, García JE. *Metodología de la investigación cualitativa*, 2a. ed. Ediciones Aljibe, Granada, España, 135-147. 1999.
- Sales CVF, Orozco MJA. Cálculo del tamaño de la muestra. En: Moreno AL, Cano VF, García RH. *Epidemiología clínica*, 2ª ed. McGraw-Hill Interamericana, México, 261-274. 1994.
- Scheaffer RL, Mendelhall W, Ott L. *Elementos de muestreo*. Grupo Editorial Iberoamérica, México, 1987.
- Schlesselman JJ. *Case-control studies. Design, conduct, analysis*. Oxford University Press, EUA, 144-170. 1982.

Silva ALC. *Muestreo para la investigación en ciencias de la salud*. Ediciones Díaz de Santos, Madrid, España, 1993.

Ulin PR, Robinson ET, Tolley EE. *Investigación aplicada en Salud Pública. Métodos cualitativos*. Publicación científica y técnica No. 614. Washington DC: OPS/OMS-Family Health International-United States Agency International Development, 48-54, 2006.



# Índice alfabético

## booksmedicos.org

Nota: el número de página en **negritas** indica cuadro, el número de página en *cursivas* indica figura.

### A

Adaptación y traducción de cuestionarios, 136  
AHRQ, clasificación del nivel de evidencia, **53**  
Ajuste de tiempo, 231  
Aleatorización, por conglomerados o grupos, 62  
    proceso de, 59  
    simple, 59  
Algunas notas de precaución, 231  
Amenidades, 262  
Análisis, de datos, 31  
    de gatillo, 227  
    de Markov, 223  
    de varianza, 32  
    estadístico, selección del, 275  
    por protocolo, 66  
    por subgrupos, 66  
Análisis de decisión, 218  
    en la práctica médica, 216  
    pasos a seguir para efectuar el, 222  
    técnicas del, 221  
Análisis de sensibilidad, 223, 225, 227  
    de una vía, 226  
    probabilístico, 228  
Analizar, 26  
Analogía, 162  
Aprender, 26  
Apuesta estándar (*standard gamble*), 229  
Árbol de decisión, construir un, 221  
Artículos, selección de los, 100  
    y bases de datos electrónicas, búsqueda de, 99  
Asignar probabilidades, 222  
Asignar utilidades, 223  
Asociación, causal, 159  
    fuerza de, 159  
    válida, 155  
Atención, de la salud basada en la evidencia, 98  
    médica, tercera revolución en la, 256  
Atención, calidad de la, criterios implícitos y explícitos, 263  
    definición de, 257  
    desarrollo de instrumentos de medición, 272  
    epidemiología, 271  
    estructura, 260  
    garantía, 268  
    incorporación de preferencias de los pacientes, 270  
    mala, 262  
    médica, 257  
        evaluación de la, 256  
    modelo unificador de, 259  
    modelos de evaluación de la, 260  
    nivel clínico, 266  
    perspectivas de la, 270

    proceso, 260  
    promoción de la investigación de la, 271  
    relación proceso-resultado, 266  
    resultado, 261  
    sistemas de salud, 265  
    tipos de estudios de, 264

Atributos de una medición, 117  
Auditoría clínica, 262  
Azar, 155

### B

Bioética y epidemiología clínica, 87  
Buenas prácticas clínicas (BPC), 70

### C

Cadenas de Markov, 224  
Calidad de vida, 245  
    aproximaciones a su medición, 243  
    concepto de, 246  
    enfoque histórico de la, 245  
    instrumentos para medir, relacionada con la salud, 249  
    investigación relacionada con la salud, 247  
    relacionada con la salud, 246  
Calidad metodológica, evaluación de la, 102  
Carta de consentimiento, 32  
Casos, definición de, 43  
Casos y controles, 26  
    diseño de un estudio de, 157  
Causa-efecto, 7  
Causalidad, determinación de la, 155  
    en medicina, concepto de, 153  
    importancia del conocimiento de la, 154  
Causas suficientes, 153  
Cegamiento, tipos y características del, **61**  
Censuramiento, 208  
Ciencia, 19  
Chi cuadrada, 32  
Clínica bioestadística, 5  
Clinimetría, 5, 109  
    antecedentes, 111  
    proceso de medición, 112  
Cocientes de probabilidad positivos y negativos, 180  
Coeficiente, de correlación, 32  
    de correlación intraclase ( $R_i$ ), 132  
    de variación, 129  
    kappa, cálculo del, 131  
Coherencia, 162  
Cohorte(s), de seguimiento, 199  
    diseño de, 198  
    diseño de un estudio de, 157  
    y experimentales, 26

Colaboración Cochrane, 106  
 Comités de Ética y Conferencia Internacional de Armonización y Buenas Prácticas Clínicas, 70  
 Concordancia, evaluación de la, 128  
 Conducción del ensayo clínico, 63  
 Conferencia Internacional de Armonización y Buenas Prácticas Clínicas, 70  
 Confiabilidad, 109, 121  
   homogeneidad y otros procedimientos para evaluar, 134  
 Confirmación o refutación de una teoría, 23  
 Conocimiento, científico, 19  
   vulgar, 19  
 Consentimiento informado, contenido de la carta de, **32**  
 Consistencia, 117, 161  
   interensayo o interobservadores, 118  
 Construcción de índices o escalas de medición, 137  
 Control(es), definición de, 44  
   selección del, 57  
 Convergencia y discriminación, 120  
 Credibilidad epidemiológica, 162  
 Criterios, de causalidad de Bradford-Hill, **160**  
   de eliminación, 28  
   de exclusión, 28, 55  
   de inclusión, 28, 55  
   explícitos, 264  
   implícitos, 263  
 Cuestionario, cómo organizar la puntuación final de un, *138*  
   de calidad de vida, 229-230  
   SF-36, 251  
 Curva de Gauss, distribución normal según, *142*  
 Curva ROC, 181  
   comparación de pruebas diagnósticas mediante, 184

## D

Datos, blandos, 111  
   duros, 111  
   extracción de los, 102  
   organización y análisis de, 31  
 Decisiones, estadísticas, 289  
   explicatorias, 110  
   terapéuticas, 110  
 Desarrollo de la Epidemiología Clínica y su método, 1  
 Descripción operativa, 116  
 Desviación estándar, 143  
 Determinación de las diferencias importantes, 67  
 Diagnóstico y evaluación de pruebas diagnósticas, 166  
 Diferencias mínimamente importantes para el clínico, 67  
 Direccionalidad, 39  
 Diseños, alternativos al ensayo clínico aleatorio y paralelo, 60  
   cruzados, 62  
   cuasi-experimentales, 36  
   descriptivos o comparativos, 40  
   epidemiológicos, **37**, **40**  
   factoriales, 61  
   longitudinales, 40  
   observacionales, 36  
   retrospectivos y prospectivos, 38  
   transversales, 40  
 Diseños de investigación, direccionalidad, 36  
   interferencia del investigador, 35

  número de mediciones, 36  
   temporalidad, 35  
 Diseños experimentales, 36  
   fases de los, 70  
   preclínica, 71  
 Diseños metodológicos, criterios utilizados en los, 35  
   en epidemiología, 35  
 Disponibilidad de una técnica adecuada, 23  
 Doxa, 19

## E

Efecto(s), adversos, 58  
   clasificación de, 82  
   del azar, 6, 7  
 Eficacia, 168  
   de una prueba, 171  
 Enfermedad, definición de, 171  
 Enmascaramiento o cegamiento, 59  
 Ensayo clínico, aleatorio, 52, 156  
   diseño de un, *157*  
   aleatorizado (ECA), 54  
   análisis del, 65  
   controlado o aleatorizado, 37  
   planeación del, 54  
 Ensayo por equivalencia, 62  
 Epidemiología Clínica, 4, 6  
   desarrollo histórico de la, **5**  
*Episteme*, 19  
 Error, aleatorio, 123  
   de medición, 122  
   de tipo I y de tipo II, ocurrencia de, **288**  
   sistemático, 123, 124  
 Escalas, análogas visuales, 229-230  
   de Apgar, *113*  
   de medición, *115*  
   variables y, 28  
   de proporción, 115  
   de razón, 115  
   para uso local, 136  
 Especificidad, 161, 170, 173, 216  
   cálculo de la, 171  
   confiabilidad de los valores de, 176  
 Estadística, 276  
   analítica, 276  
   descriptiva, 276  
 Estándar diagnóstico ideal, 169  
 Estimación del pronóstico de la enfermedad, 198  
 Estímulo, 228  
 Estudio experimental, comparativo, 26  
   longitudinal, 26  
   prospectivo, 26  
 Estudio observacional, 26  
   comparativo o descriptivo, 26  
   retrospectivo o prospectivo, 26  
   transversal o longitudinal, 26  
 Estudios, de aplicación, 36  
   de casos y controles, 42  
   ventajas y limitaciones de, **46**  
   de cohorte, 45  
   ventajas y limitaciones de, **47**  
   de corte transversal, 40  
   de intervención, 37

- de medición simultánea, 39
- epidemiológicos específicos, 42
- etiológicos, 41
- experimentales, 37, 52
- exploratorios, 41
- Framingham, 4
- instantáneos, 40
- progresivos, 39
- prospectivos, 38
- regresivos, 39
- retrospectivos, 38
- seguimiento y apego al protocolo de, 63
- seudoaleatorizados, 62
- transversales, 26, 40
- verticales, 40
- Evento, adverso, 76
  - centinela, 265
- Evidencia experimental, 162
- Explicaciones, contradictorias, 12
  - tentativas, 12
- Explosión farmacológica, 73
- F**
- Factor, de riesgo, 199-200
  - pronóstico, 199-200
- Falsos, negativos, 182
  - positivos, 182
- Falla, 208
- Farmacovigilancia, 73
  - a nivel internacional, 76
  - antecedentes, 74
  - características de un sistema de, 78
  - ¿Cómo funciona un sistema de?, 78
  - definición, 76
  - en la práctica clínica, 83
  - en México, 84
    - flujo de información del proceso de, 85
  - eventos relevantes que dieron origen a la, 75
  - funciones del sistema de, 80
  - medidas de seguridad, 83
  - notificación individual de un caso, 82
- Función acumulada de sobrevida, 208
- Fundamentación, 20
  - antecedentes, 20
  - hipótesis, 24
  - justificación, 24
  - marco teórico, 24
  - objetivos, 25
  - planteamiento del problema, 22
  - pregunta de investigación, 22
  - título, 20
- G**
- Garantía o aseguramiento de la calidad, 268
- Gradiente dosis-respuesta, 161
- Grados de libertad (gl), 286
- Gráfica de Gant para programación logística de un proyecto de investigación, 33
- Gráfico, de árbol Forest Plot, 103
  - de embudo (*funnel plot*), 104
- Grupo, selección del, expuesto, 47
  - no expuesto, 48

- H**
- Heterogeneidad, modelos de análisis  $\chi^2$ , 103
- Heurística, de ancla y ajuste, 217
  - de memoria, 217
  - de representatividad, 217
- Hipótesis, científicas, 13
  - y problema de investigación, 12
- Homogeneidad y otros procedimientos para evaluar confiabilidad, 134

- I**
- Implementación, 32
- Indicador, 114
- Individuos en estudio, selección de, 280
- Información, presentación de la, 31
  - recomendaciones adicionales para obtener, 138
- Instrumentos, de medición, 115
  - específicos o funcionales, 250
- Instrumentos genéricos, 249
  - medidas de pregunta única, 250
  - medidas de utilidad o preferencia, 250
  - perfiles de salud, 250
- Interpretación, de los resultados, 67
  - e integración de la información, 228
- Interpretar, 26
- Intervalos de confianza, 296
- Intervención, selección de la, 57
- Investigación, aspectos éticos de la, 69
  - científica, 19, 23
  - clínica, 2
  - clínica sea ética, requisitos para que una, 90
  - con riesgo mayor que el mínimo, 88
  - con riesgo mínimo, 88
  - cuantitativa, 15
  - epidemiológica, 3
  - lineamientos para el diseño del proyecto de, 19
  - sin riesgo, 88

- J**
- Juicio, 229

- L**
- Laguna objetiva del conocimiento, 11
- Ley General de Salud, 87
- Lineamientos éticos, 32
- Logística, 33

- M**
- Malclasificación no diferencial y diferencial, 125
- Marcadores intermedios, 58
- Marco, muestral, 293
  - teórico, función del, 24
- Medicamento, etapas para el desarrollo de un, 74
- Medición, concepto de, 31
  - fases en el proceso de, 31
- Metaanálisis, en la medicina moderna, 97
  - en medicina, 94
  - resumen numérico de los estudios, 102
- Método, científico, 1, 12
  - de Bland y Altman, 133

- Método, de Kaplan-Meier, fórmula para estimar la función acumulada de sobrevida por el, 210  
 para el cálculo de  $\hat{S}_t$ , 209  
 de medición, 116  
 de selección, 45  
 directo e indirecto, 29  
 para obtener calidad por inspección, 269
- Metodología, diseño del estudio, 26  
 población o muestra, 26
- Modelos, de garantía de calidad, 268  
 probabilísticos, 6
- Monitoreo, 64
- Mortalidad, 213
- Movimiento de los indicadores sociales, 245
- Muestra, 27, 294  
 métodos de selección de la, 28  
 tamaño de, fórmulas para el cálculo de un, **305**  
 insuficiente, 303
- Muestreo, 297  
 aleatorio simple, 298  
 estratificado, 299  
 no probabilístico, 299  
 por conglomerados, 299  
 por conveniencia, 300  
 probabilístico, condiciones de aplicación, 298  
 sistemático, 299  
 y cálculo de tamaño de muestra, 292
- N**
- Negociar con el tiempo (*time trade-off*), 229
- Nivel de significancia, 302
- No., casos, 43
- Nodo, de decisión, 221  
 de probabilidad, 221  
 terminal, 222
- Normalidad, comparación de los criterios de, **151**  
 concepto aislado de, 141  
 concepto relacionado de, 146  
 definición correlacionada (multivariada), 140  
 disponibilidad de una terapéutica, 150  
 en medicina: usos y limitaciones, 140  
 exposición a un factor de riesgo, 148  
 presencia de síntomas y/o signos, 146
- Nueva unidad de análisis, 103
- Número de variables, 280
- O**
- Objetivos, del estudio, 279  
 factores indispensables para definir los, **26**
- Observaciones, censuradas, 208  
 no censuradas, 209
- Observador, 117
- Obtención de la información, reglas generales en el proceso de, 30
- Ordenar, 26
- Organización y análisis de datos, 31
- P**
- Placebo, 57
- Plausibilidad o credibilidad biológica, 162
- Población, accesible, 294  
 blanco, 293  
 de estudio, selección de la, 54  
 muestra, ejemplo de gráfica para determinar una, 27  
 o universo, 293
- Posible, probable o frecuente, 220
- Predicción causal, 13
- Pregunta, de investigación, 11, 54  
 principal, 54  
 secundaria, 54
- Principio, analítico por intención de tratar, 65  
 de autonomía, 69  
 de justicia, 69  
 de no maleficencia, 69
- Probabilidad, 6  
 de ocurrencia, 7  
 diferentes distribuciones de, 226
- Problema(s), científico en medicina, 13  
 ejemplo del origen de un, 16  
 prácticos, 22  
 teóricos, 22  
 y objetivos, definición del, 98
- Problema(s) de investigación, 10  
 fundamentación del, 10  
 hipótesis y, 12  
 requisitos de un, 14
- Proceso de Markov, 224  
 ejemplo de, 236  
 representación esquemática de, 224
- Proceso de medición, componentes del, 113
- Programa Internacional de Farmacovigilancia de la OMS, 77
- Programas de detección, normas generales para la aplicación de los, **187**
- Pronóstico, aplicación del análisis de sobrevida en estudios de, 208  
 medición de eventos incidentes de, 200  
 medición de la fuerza de asociación en estudios de, 205  
 medición del, 198
- Protocolo, 19  
 de investigación, componentes del, 20, **21**  
 y proyecto, 19
- Proxyindicador, 114
- Prueba(s),  $\alpha$  de Cronbach, 134  
 de detección temprana, 186  
 de heterogeneidad, 104  
 en paralelo, 189  
 en serie, 188  
 kappa ( $\kappa$ ), 129  
 kappa ponderada ( $\kappa_w$ ), 131  
 múltiples, 188  
 potencia de la, 279
- Prueba(s) diagnóstica, características de la población, 185  
 condiciones necesarias para la validación de una, 185  
 precisión de los resultados, 186  
 presentación de resultados indeterminados, 186  
 reproducibilidad de la prueba, 186  
 sesgo de medición, 186  
 sesgo de selección, 186  
 subgrupos adecuados, 185
- Prueba(s) estadística, lista de, **278**  
 selección de la, 281
- Prueba(s) estadística, etapas de la aplicación de, 285  
 definir la prueba estadística a utilizar, 285

- grados de libertad (gl), 286
  - identificar el valor crítico, 287
  - naturaleza de datos, 285
  - nivel de significancia, 286
  - planteamiento de las hipótesis estadísticas, 285
  - regla de decisión, 287
  - Psicometría, 109
  - Punto, de arranque, 299
    - fundamental, 65
- R**
- Razón(es), 180
    - de densidades de incidencia, 213
    - de densidades de mortalidad (RDM), 207
    - de verosimilitud, 180
      - negativa, 180
      - positiva, 180
  - Razonamiento clínico, 5
  - Reacción adversa a medicamentos (RAM), 76
    - estrategias o métodos de recolección de información sobre, 79
  - Realidad y resultado de la prueba, relación entre, 303
  - Recolección de información, procedimientos de, 29
  - Recomendaciones adicionales para obtener información, 138
  - Recursos, financieros, 33
    - humanos, 33
    - materiales, 33
    - y logística, 32
  - Referencias bibliográficas, 34
  - Regla de decisión, 287
  - Reglamento en materia de investigación para la salud, 87
  - Resolución del escenario inicial, 252
  - Respuesta, forma de, 229
  - Resultado(s), características psicológicas del, 261
    - del ensayo, importancia y validez de, 67
  - Revisiones narrativas *vs.* revisiones sistemáticas, 95
  - Revisiones sistemáticas, 94
    - evaluación de, 106
    - importancia de las, 97
    - utilidad de las, 98
  - Riesgo relativo (RR), 205, 213
- S**
- Seguimiento, 49
  - Sensibilidad, 170, 173, 216
    - cálculo de la, 171
    - confiabilidad de los valores de, 176
  - Serie de tiempo, 62
  - Sesgo, 123, 124, 156
    - control del, 127, 212
    - de confusión, 127
    - de información, 126
    - de selección, 125
    - en estudios de pronóstico, 211
    - tipos de, 125
  - Si ocurre X, sucederá Y, 13
  - Significancia, clínica, 283
    - estadística, 283
  - Sistemas de salud, 265
  - Solución, identifique el problema y las alternativas de, 232
  - Sujeto de la medición, 117
  - Suma de tiempos en riesgos (STR), 206
- T**
- t* de Student, 32
  - Tabla de contingencia de  $2 \times 2$ , 160
  - Tamaño del efecto (TE), 102
  - Tamiz o escrutinio, 186
  - Técnicas econométricas, 229
  - Temporalidad, 161
  - Teoría, de la comprensión de la enfermedad, 3
    - de la sensibilidad, 135
  - Testigos, 43
    - definición de, 44
  - Tiempo, de falla, 208
    - en riesgo, 200
    - no censurado, 211
  - Tomar la decisión, 223
- U**
- Unidad de análisis, 36
  - Uppsala Monitoring Centre* (UMC), 77
  - Utilidad, 228
    - esperada, 223
    - técnicas de valuación de, 229-230
  - Utilitarismo, 245
- V**
- Validez, 118, 168
    - de apariencia, 119
    - de constructo, 120
    - de contenido, 119
    - de criterio, 120
    - de expresión, 119
    - externa, 119
    - interna, 119
  - Valor(es), cuantitativos, 229
    - de predicción de una prueba negativa (VP-), 177
    - de predicción de una prueba positiva (VP+), 177
    - predictivos, 177
  - Variabilidad y error en la medición, 122
  - Variables, 113
    - clasificación de, 277
    - clasificación y características de las, 114
    - de desenlace, 58
    - multidimensionales, 114
    - nominales, 114
    - ordinales, 114
    - subrogadas, 58
    - unidimensionales, 114
    - y escalas de medición, 28
  - Variables cualitativas, 28, 114
    - nominales, 114
    - ordinales, 114
  - Variables cuantitativas, continuas, 29, 114
    - discretas, 29, 114
  - Variación, coeficiente de, 129
  - Vigilancia, farmacéutica internacional, 77
    - poscomercialización, 83