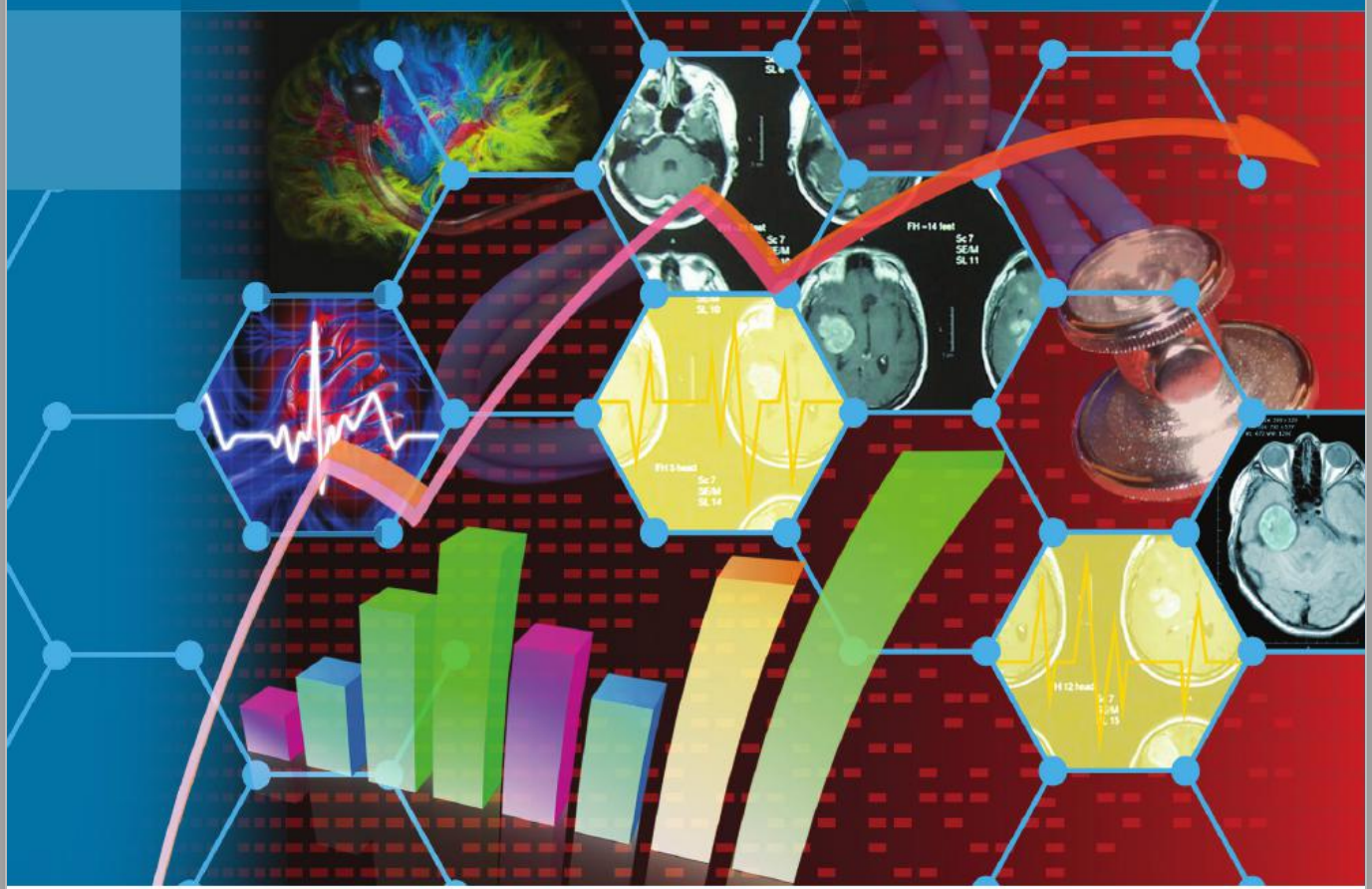
Álvaro J. Ruiz Morales • Carlos Gómez-Restrepo

Epidemiología Clínica

Investigación Clínica Aplicada

2.ª EDICIÓN



EDITORIAL MEDICA panamericana

Epidemiología Clínica

Investigación Clínica Aplicada

2.ª Edición

Epidemiología Clínica

Investigación Clínica Aplicada

2.ª Edición

Álvaro J. Ruiz Morales

Médico. Especialización en Medicina Interna y en Hipertensión Arterial. Maestría en Epidemiología Clínica. Profesor titular de Medicina Interna y de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.

Carlos Gómez-Restrepo

Médico. Especialización en Psiquiatría. Psicoanalista. Maestría en Epidemiología Clínica. Profesor titular de Psiquiatría y de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.

Libros Medicos PDF

EDITORIAL MEDICA panamericana

BOGOTÁ • BUENOS AIRES • CARACAS • MADRID • MÉXICO • PORTO ALEGRE

Visite nuestra página web:
www.medicapanamericana.com

Gracias por comprar el original. Este libro es producto del esfuerzo de profesionales como usted, o de sus profesores, si usted es estudiante. Tenga en cuenta que fotocopiarlo es una falta de respeto hacia ellos y un robo de sus derechos intelectuales.

Las ciencias de la salud están en permanente cambio. A medida que las nuevas investigaciones y la experiencia clínica amplían nuestro conocimiento, se requieren modificaciones en las modalidades terapéuticas y en los tratamientos farmacológicos. Los autores de esta obra han verificado toda la información con fuentes confiables para asegurarse de que esta sea completa y acorde con los estándares aceptados en el momento de la publicación. Sin embargo, en vista de la posibilidad de un error humano o de cambios en las ciencias de la salud, ni los autores, ni la editorial o cualquier otra persona implicada en la preparación o la publicación de este trabajo, garantizan que la totalidad de la información aquí contenida sea exacta o completa y no se responsabilizan por errores u omisiones o por los resultados obtenidos del uso de esta información. Se aconseja a los lectores confirmarla con otras fuentes. Por ejemplo, y en particular, se recomienda a los lectores revisar el prospecto de cada fármaco que planean administrar para cerciorarse de que la información contenida en este libro sea correcta y que no se hayan producido cambios en las dosis sugeridas o en las contraindicaciones para su administración. Esta recomendación cobra especial importancia con relación a fármacos nuevos o de uso infrecuente.

Los editores han hecho todos los esfuerzos para localizar a los poseedores del ***copyright*** del material fuente utilizado. Si inadvertidamente hubieran omitido alguno, con gusto harán los arreglos necesarios en la primera oportunidad que se les presente para tal fin.



ISBN 978-958-8443-58-4 Versión electrónica

EDITORIAL MEDICA panamericana

Visite nuestra página web:
http://www.medicapanamericana.com

Todos los derechos reservados.
Este libro o cualquiera de sus partes no podrán ser reproducidos ni archivados en sistemas recuperables, ni transmitidos en ninguna forma o por ningún medio, ya sean mecánicos o electrónicos, fotocopiadoras, grabaciones o cualquier otro, sin el permiso previo de Editorial Médica Internacional Ltda.

© 2015 EDITORIAL MÉDICA
INTERNACIONAL LTDA.
Cra. 7ª A N° 69-19
Bogotá, D. C., Colombia

Corrección de estilo: Gustavo Patiño Díaz

Diagramación: Yolanda Velásquez L.

Diseño de carátula: Elkin Pimienta

Índice analítico: Gustavo Patiño Díaz

ARGENTINA
Marcelo T. de Alvear 2145 (C1122AAG)
Ciudad Autónoma de Buenos Aires
República Argentina
Tel.: (54-11) 4821-2066 / 5520
Fax: (54-11) 4821-1214
info@medicapanamericana.com

COLOMBIA
Carrera 7ª A Nº 69-19
Bogotá, D. C.
Tel.: (57-1) 345-4508 / 314-5014
Fax: (57-1) 314-5015 / 345-0019
infomp@medicapanamericana.com.co

ESPAÑA
Quintanapalla, 8. 4ª Planta 28050 Madrid.
Madrid
Tel.: (34) 91-1317800
Fax: (34) 91-1317805
info@medicapanamericana.es

MÉXICO
Hegel N° 141, 2° piso
Colonia Chapultepec Morales
Delegación Miguel Hidalgo
México D. F.
C.P. 11570
Tel.: (5255)5250-0664 5250-9470
Fax (5255) 2624-2827
infomp@medicapanamericana.com.mx

VENEZUELA
Edif. Polar, Torre Oeste, Piso 6, Ofic. 6C
Plaza Venezuela, Urbanización los
Caobos, Parroquia El Recreo,
Municipio Libertador
Caracas Distrito Capital
Tel.: (58-0212)793-6906/1666/5985
Fax (58-0212) 793-5885
infomp@medicapanamericana.com.ve

Colaboradores

Jhon Camacho

Ingeniero Electrónico. Maestría en Informática Biomédica. Profesor instructor. Departamento de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 30, 31, 32

Ana María de la Hoz Bradford

Médica. Maestría en Epidemiología Clínica. Bogotá, D. C., Colombia.
Caps. 11, 22

Martha Beatriz Delgado Ramírez

Médica. Especialización en Anestesiología. Maestría en Epidemiología Clínica. Profesora asociada de Anestesiología y de Epidemiología Clínica. Directora de Carrera de Medicina, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Cap. 26

Rodolfo José Dennis Verano

Médico. Especialización en Medicina Interna y Neumología. Maestría en Epidemiología Clínica. Profesor titular de Medicina Interna y de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Director, Departamentos de Medicina Interna y de Investigaciones, Fundación Cardioinfantil - Instituto de Cardiología. Bogotá, D. C., Colombia.
Cap. 24

Andrés Duarte Osorio

Médico. Especialización en Medicina Familiar. Maestría en Epidemiología Clínica. Profesor asociado de Medicina Familiar y Medicina Preventiva, Facultad de Medicina, Pontificia Universidad Javeriana. Médico familiar, Hospital Universitario San Ignacio. Bogotá, D. C., Colombia.
Cap. 14

Carlos Gómez-Restrepo

Médico. Especialización en Psiquiatría. Psicoanalista. Maestría en Epidemiología Clínica. Profesor titular de Psiquiatría y de Epidemiología Clínica. Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. (8, 13, 16, 19, 28)

Claudia Marcela Granados Rugeles

Médica. Especialización en Pediatría. Maestría en Epidemiología Clínica. Profesora asistente de

Pediatría y de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Cap. 27

Luis Fernando Jaramillo García

Médico. Especialización en Patología. Maestría en Epidemiología Clínica. Profesor asociado, Departamento de Patología, Facultad de Medicina. Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Cap. 12

Juan Manuel Lozano León

Médico. Especialización en Pediatría. Maestría en Epidemiología Clínica. Profesor titular, Departamento de Investigación Médica y en Ciencias de la Salud, Colegio de Medicina, Universidad Internacional de la Florida, Miami, FL, Estados Unidos.
Caps. 1, 9, 15, 20, 35

Socorro Moreno Luna

Psicóloga. Maestría en Epidemiología Clínica. Profesora asistente de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Cap. 7

Óscar Mauricio Muñoz Velandia

Médico. Especialización en Medicina Interna. Maestría en Epidemiología Clínica. Profesor asistente de Medicina Interna, Departamento de Medicina Interna, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 3, 22

María Beatriz Ospina García

Psicóloga. Maestría y doctorado en Epidemiología. Directora Científica Adjunta, Respiratory Health Strategic Clinical. Alberta, Canadá.
Caps. 8, 19, 28

Hoover Quitián Reyes

Economista. Magíster en Investigación en Ciencias Económicas. Docente, Economía de la Salud, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 21, 23

María Nelcy Rodríguez Malagón

Estadística. Maestría en Salud Pública en Bioestadística. Profesora asociada, Bioestadística básica y avanzada, Departamento de Epidemiología Clínica y Bioestadística, Pontificia Universidad Javeriana. Profesora asociada, Departamento de Estadística, Universidad Nacional de Colombia. Bogotá, D. C., Colombia.
Caps. 5, 18, 29

María Ximena Rojas Reyes

Enfermera. Maestría en Epidemiología Clínica. Maestría en Salud Pública. Doctorada en Metodología de Investigación Biomédica. Profesora asociada de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 20, 25, 26

Martín Alonso Rondón Sepúlveda

Estadístico. Maestría en Bioestadística. Profesor Asociado e Investigador, Bioestadística, Departamento de Epidemiología Clínica y Bioestadística, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 6, 8, 33

Álvaro J. Ruiz Morales

Médico. Especialización en Medicina Interna y en Hipertensión Arterial. Maestría en Epidemiología Clínica. Profesor titular de Medicina Interna y de Epidemiología Clínica, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Caps. 3, 10, 13, 27, 34, 36, 37

Juan Gabriel Ruiz Peláez

Médico. Especialización en Pediatría. Maestría en Epidemiología Clínica. Profesor titular de Pediatría y de Epidemiología Clínica, Pontificia Universidad Javeriana. Pediatra, Hospital Universitario San Ignacio. Bogotá, D. C., Colombia.
Caps. 2, 4, 5, 15, 17, 29

Jenny Alexandra Severiche Báez

Administradora de Empresas. Maestría en Educación. Asistente Decanatura Académica. Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.
Cap. 11

Fernando Suárez Obando

Médico. Especialización en Genética Médica. Maestría en Epidemiología Clínica. Maestría en Informática Médica. Profesor Asistente de Genética Humana, Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.

Caps. 36, 38

Luis Felipe Uriza Carrasco

Médico. Especialización en Radiología. Profesor asociado, Facultad de Medicina, Pontificia Universidad Javeriana. Bogotá, D. C., Colombia.

Cap. 18

Prólogo

Esta obra nació en el seno del Departamento de Epidemiología Clínica y Bioestadística de la Pontificia Universidad Javeriana, pionero en el país de este método científico. Fue el Dr. Arturo Morillo Quiñones (QEPD), decano en su momento de la Facultad de Medicina, quien gracias a su tesón y esfuerzo logró para el Departamento de Epidemiología el apoyo de la Fundación Rockefeller, de International Clinical Epidemiology Network (Inclen) y de la misma Universidad, para formar en el exterior un grupo de profesores en programas reconocidos.

A su regreso, estos profesores le dieron vida a la Unidad de Epidemiología Clínica, que creció y se fortaleció hasta llegar a ser hoy en día un sólido departamento, que ha alcanzado el reconocimiento como 'Centro de Excelencia en Investigación' por Colciencias, con un grupo de investigación A1, conformado por sus destacados profesores e investigadores. Su trabajo les ha permitido tener un programa de Maestría en Epidemiología Clínica muy bien posicionado, complementado ahora con un programa de Doctorado en Epidemiología Clínica y con una Maestría en Bioestadística, una de las grandes fortalezas del grupo en este campo de formación.

La epidemiología clínica ha entrado a formar parte de los currículos de la medicina en las últimas décadas y ha permitido introducir a los estudiantes en la importancia del método científico, de la rigurosidad científica y del interés por buscar la mejor evidencia; así, se ha convertido en una herramienta sólida para la búsqueda de las respuestas a preguntas clínicas en pro del mejor cuidado de los pacientes.

La epidemiología clínica aporta a todos los que de alguna manera tienen que ver o inciden desde su profesión en el cuidado de las personas y en el cuidado de la salud. Para el médico en general, independientemente del área en la cual se desempeñe, bien como clínicos, administradores, gestores de política pública, actores del sistema de salud, formadores de nuevas generaciones o investigadores, la epidemiología clínica se ha convertido en una aliada estratégica en el trabajo del día a día.

Este libro está distribuido en 38 capítulos que incluyen varios temas que no formaban parte de la edición anterior; los restantes traen actualizaciones, avances e innovaciones surgidas desde la primera publicación. Los temas se organizan desde los fundamentos básicos, la pregunta de investigación y el muestreo, y avanzan a través

de los diferentes tipos de diseños, la conducción del estudio, el análisis y manejo de los datos, su presentación y aspectos éticos y de autoría.

El lenguaje, la organización, la presentación y profundidad de los contenidos, así como su calidad harán de la obra una herramienta valiosa y un texto de consulta en diferentes niveles y será una contribución muy importante a la investigación en Hispanoamérica y a la formación de investigadores, pero también tendrá un valor importante en la aplicación de este método científico, en el trabajo clínico al lado del paciente, en el diseño de políticas públicas alrededor de la mejor evidencia científica y en la generación de nuevo conocimiento.

Esta nueva edición de ***Epidemiología clínica*** es el producto del trabajo de los profesores del Departamento y de sus egresados, que han contribuido a la construcción de este libro con la solidez y rigurosidad que los caracteriza, con sus conocimientos y experiencias, lo que asegura la calidad de la obra y el aporte que hará a la salud.

Mary Bermúdez Gómez, MD, MSc

Decana Académica, Facultad de Medicina
Profesora titular de Medicina Interna
y de Neumología
Neumóloga y magíster en
Epidemiología Clínica
Pontificia Universidad Javeriana-Hospital
Universitario San Ignacio

Introducción

La medicina es arte y es ciencia. Es una mezcla de datos sólidos provenientes de investigación, de experiencias obtenidas en observaciones repetidas, de acuerdos entre expertos, de prejuicios, anhelos e intereses, de habilidades y de actitudes. La calidad del ejercicio profesional reside en la ciencia y en el arte que haya en cada componente.

La toma de decisiones requiere conocimientos sólidos, destrezas y una actitud inteligente para que el resultado sea efectivo. Para esto, el profesional de la salud debe estar preparado en los conceptos básicos de la epidemiología clínica; debe conocer los principios de la medición y las diferentes clases de error; debe estar familiarizado con los diseños más utilizados en investigación clínica, con los elementos más importantes de las pruebas diagnósticas y con los fundamentos del análisis de decisiones.

Pero el conocimiento no lo es todo, como tampoco lo es la presencia de destrezas y habilidades. El componente de arte —que agrega humanismo, que personaliza el ejercicio, que tiene en cuenta los aspectos individuales y colectivos de la enfermedad— debe estar marcado por una actitud de curiosidad permanente y de crítica constante.

La palabra ***investigar*** viene del verbo latino *invenire*, que significa 'encontrar'. Y encuentra el que busca. La investigación es el resultado de la curiosidad, del afán de buscar cosas nuevas, del ejercicio de hacerse preguntas para resolver problemas, o para buscar soluciones mejores o diferentes a las existentes. Y la actitud de búsqueda la tiene quien ejercita su actitud crítica ante cada hecho o situación que se le presente.

Para aportar arte al ejercicio de la medicina y de las ciencias de la salud es indispensable el uso del espíritu crítico, la evaluación permanente de cuanto se observa y encuentra. De esta necesidad surgió el enfoque llamado ***medicina basada en evidencia***, que busca hacer óptimo el proceso de toma de decisiones mediante el uso de información sólida, que provenga de investigación de la mejor calidad.

El ejercicio profesional responsable necesita, entonces, conocimientos, habilidades y actitudes para la generación de investigación de buena calidad. Y necesita también los mismos elementos para la evaluación crítica de la investigación que se le presente.

Este libro busca ofrecer al profesional de la salud los elementos necesarios para emprender

la fascinante tarea de investigar, de producir conocimiento generalizable. Y también quiere ofrecerle los elementos suficientes para hacer una buena apreciación crítica de la literatura científica.

La producción de conocimiento y la evaluación crítica son las bases para la toma de decisiones en la atención en salud, que debe buscar el bienestar y la salud de la población, fin último de la medicina.

Contenido

Colaboradores vii

Prólogo xi

Introducción xiii

SECCIÓN 1 Definición de la pregunta de investigación

Capítulo 1 Investigación en salud: qué es y para qué sirve 3
Juan Manuel Lozano León

Capítulo 2 La pregunta de investigación 11
Juan Gabriel Ruiz Peláez

Capítulo 3 Búsqueda de la literatura médica 30
Álvaro J. Ruiz Morales • Óscar Mauricio Muñoz Velandia

SECCIÓN 2 Muestreo, medición y variabilidad

Capítulo 4 Variables y fuentes de error en la medición. Confusión, sesgo, azar 45
Juan Gabriel Ruiz Peláez

Capítulo 5 Población y muestra 61
Juan Gabriel Ruiz Peláez • María Nelcy Rodríguez Malagón

Capítulo 6 Cálculo del tamaño de muestra 76
Martín Alonso Rondón Sepúlveda

Capítulo 7 Variables no observables en epidemiología clínica: el uso de constructos .. 91
Socorro Moreno Luna

Capítulo 8 Adaptación cultural y validación de escalas 95
Carlos Gómez-Restrepo • Martín Alonso Rondón Sepúlveda
María Beatriz Ospina García

Capítulo 9 Medidas de frecuencia, de asociación y de impacto 113
Juan Manuel Lozano León

Capítulo 10 Uso de pruebas diagnósticas en medicina clínica 129
Álvaro J. Ruiz Morales

SECCIÓN 3 Diseños de investigación clínica. Estudios primarios

Capítulo 11 La propuesta de investigación 147
Ana María de la Hoz Bradford • Jenny Alexandra Severiche Báez

Capítulo 12 El protocolo de investigación. Una introducción 162
Luis Fernando Jaramillo García

Capítulo 13 Introducción a los diseños de investigación: selección del diseño apropiado 168
Álvaro J. Ruiz Morales • Carlos Gómez-Restrepo

Capítulo 14 Estudios de corte transversal 188
Andrés Duarte Osorio

Capítulo 15 El experimento clínico 198
Juan Manuel Lozano León • Juan Gabriel Ruiz Peláez

Capítulo 16 Estudio de casos y controles 226
Carlos Gómez-Restrepo

Capítulo 17 Estudios de cohorte analítica 244
Juan Gabriel Ruiz Peláez

Capítulo 18 Estudios sobre correlación y concordancia 268
Luis Felipe Uriza Carrasco • María Nelcy Rodríguez Malagón

Capítulo 19 Estudios cualitativos. 281
María Beatriz Ospina García • Carlos Gómez-Restrepo

SECCIÓN 4 Diseños de investigación clínica. Estudios secundarios o integrativos

Capítulo 20 Revisiones sistemáticas y metaanálisis de la literatura 299
María Ximena Rojas Reyes • Juan Manuel Lozano León

Capítulo 21 Introducción al análisis de decisiones 325
Hoover Quitián Reyes

Capítulo 22 Desarrollo de guías de práctica clínica. 331
Óscar Mauricio Muñoz Velandia • Ana María de la Hoz Bradford

SECCIÓN 5 Análisis económicos y evaluación de la tecnología

Capítulo 23 Introducción a la evaluación económica de tecnologías en salud 349
Hoover Quitián Reyes

Capítulo 24 Estudios de farmacoepidemiología y farmacoeconomía 367
Rodolfo José Dennis Verano

Capítulo 25 Evaluaciones de tecnología 386
María Ximena Rojas Reyes

SECCIÓN 6 Conducción de estudios

Capítulo 26 Conducción y monitorización de estudios clínicos 411
Martha Beatriz Delgado Ramírez • María Ximena Rojas Reyes

Capítulo 27 Adherencia 434
Álvaro J. Ruiz Morales • Claudia Marcela Granados Rugeles

SECCIÓN 7 Recolección, análisis y presentación de datos

Capítulo 28 Desarrollo de cuestionarios 443
Carlos Gómez-Restrepo • María Beatriz Ospina García

Capítulo 29 Plan de análisis 454
María Nelcy Rodríguez Malagón • Juan Gabriel Ruiz Peláez

Capítulo 30 Modelado 464
Jhon Camacho

Capítulo 31 Construcción de bases de datos 481
Jhon Camacho

Capítulo 32 Preparación de datos 492
Jhon Camacho

Capítulo 33 Presentación estadística de datos 502
Martín Alonso Rondón Sepúlveda

Capítulo 34 Medicina basada en evidencia 529
Álvaro J. Ruiz Morales

Capítulo 35 Apreciación crítica de la literatura 540
Juan Manuel Lozano León

Capítulo 36 Principios éticos en investigación clínica y epidemiológica 552
Álvaro J. Ruiz Morales • Fernando Suárez Obando

Capítulo 37 Autoría 572
Álvaro J. Ruiz Morales

Capítulo 38 Informática biomédica y epidemiología clínica 588
Fernando Suárez Obando

Índice analítico 599



SECCIÓN

1

Definición de la pregunta de investigación

CAPÍTULO 1 Investigación en salud: qué es y para qué sirve
Juan Manuel Lozano León

CAPÍTULO 2 La pregunta de investigación
Juan Gabriel Ruiz Peláez

CAPÍTULO 3 Búsqueda de la literatura médica
Álvaro J. Ruiz Morales, Óscar Mauricio Muñoz Velandia



CAPÍTULO 1

Investigación en salud: qué es y para qué sirve

JUAN MANUEL LOZANO LEÓN

INTRODUCCIÓN

Este libro se centra en una de las principales aplicaciones de la epidemiología clínica: los principios para conducir la investigación clínica o epidemiológica, de modo que sea relevante y de alta calidad. En este primer capítulo se ofrece una definición práctica de la investigación en salud y de sus propósitos, así como una descripción de las diversas formas que los humanos emplean para obtener conocimientos novedosos, que les permitan comprender y explicar la realidad que los rodea.

DEFINICIÓN DE INVESTIGACIÓN EN SALUD

Para la Comisión de Investigación en Salud para el Desarrollo, la investigación en salud es:

> "La generación de conocimiento nuevo, producido a través del método científico, para identificar y controlar los problemas de salud"[1].

Es conveniente resaltar varios aspectos de esta definición. En primer lugar, la investigación en salud es un proceso que busca generar conocimientos ***novedosos***; **es decir, aumentar la** comprensión de la salud y de la enfermedad. En segundo término, propone que este proceso se haga empleando una estrategia específica para generar el conocimiento; en concreto, el método científico. Como se discutirá más adelante en este mismo capítulo, hay otras formas diferentes del método científico para generar conocimiento. Por último, señala un propósito concreto para la investigación: el conocimiento generado debe permitir no solo comprender mejor los fenómenos de la salud y de la enfermedad, sino que dicho conocimiento debe contribuir a la solución de los problemas en estos campos.

¿POR QUÉ SE DEBE HACER INVESTIGACIÓN EN SALUD?

El conocimiento generado a través de la investigación puede contribuir de varias formas a solucionar los problemas de salud. A través de la investigación, se pueden identificar y comprender los principales problemas en el campo. Sin embargo, la tarea de identificar los principales problemas en salud no debe verse como algo relevante solo para las enfermedades nuevas. Existe un renovado interés en identifi-

car aspectos tan básicos como la frecuencia; en particular, la de entidades como la tuberculosis y la malaria, que han afectado a la humanidad por siglos.

EJEMPLO

La investigación permitió conocer, relativamente en poco tiempo, muchos aspectos del síndrome respiratorio agudo severo (SARS, por sus iniciales en inglés), una enfermedad nueva aparecida en China en 2002. Transcurridos pocos meses desde el comienzo de la epidemia, se identificó su agente causal, así como algunos de los mecanismos de transmisión y un posible reservorio animal, y se desarrollaron pruebas para el diagnóstico de casos sospechosos.

La investigación también es fundamental para el desarrollo de intervenciones o de recursos tecnológicos útiles para el mantenimiento de la salud, para la prevención de la enfermedad, para su diagnóstico oportuno y para el tratamiento de los afectados. Con referencia a estas intervenciones y a la tecnología asociada, es necesario señalar que la investigación no se limita a su desarrollo. La investigación también es imprescindible para orientar y acelerar su aplicación. Este es un campo de particular pertinencia para los investigadores de regiones menos desarrolladas, que a menudo deben emplear los recursos producidos en otros contextos para solucionar sus propios problemas.

Dado que la mayor parte de la investigación en salud se lleva a cabo en los países más desarrollados, pueden surgir interrogantes respecto a la efectividad de las intervenciones desarrolladas con base en dicho conocimiento cuando se aplican en otras circunstancias geográficas, demográficas, biológicas, sociales, etc. De ahí se desprende que la investigación en salud no es un campo reservado para los científicos que trabajan en el laboratorio: en ella también se requieren conocimientos generados por epidemiólogos, clínicos, estadísticos, científicos sociales y economistas, entre otros.

Finalmente, la investigación en salud permite evaluar los problemas en el campo a lo largo del tiempo, así como sus modificaciones ante la implementación de intervenciones para controlarlos.

Lo descrito en los párrafos anteriores podría resumirse diciendo que el propósito final de la investigación en salud es desarrollar una base de conocimiento que permita la mayor efectividad posible de la práctica profesional en el área. El concepto de la práctica en salud basada en evidencia representa este principio: los cuidados para mantener y recuperar la salud de los individuos dependen de nuestra habilidad para tomar decisiones que tengan bases en un conocimiento científicamente sólido, basado en la mejor evidencia disponible en el momento.

FORMAS DE ADQUIRIR EL CONOCIMIENTO

Existen muchas estrategias a través de las cuales se puede obtener nuevo conocimiento[2-5]:

- **Tradición:** Las culturas aceptan ciertos conceptos o verdades tal como se les dan. En otras palabras, se acepta que son veraces simplemente porque siempre se ha considerado que son ciertas, sin necesidad de validación externa. Sin embargo, como fuente del conocimiento, la tradición puede tener serias limitaciones en el campo de la salud, no solo debido a que, por definición, no se haya probado la validez de sus postulados, sino porque no se han comparado sus propuestas con otras que pueden ser mejores alternativas. Adicionalmente, el excesivo énfasis en el conocimiento tradicional puede impedir la búsqueda de nueva información.

EJEMPLO

Por décadas, se consideró que la única forma de tratar la apendicitis aguda era la apendicetomía. La evidencia reciente sugiere que los casos de apendicitis no complicada podrían tratarse con antibióticos y sin cirugía, y que el tratamiento no quirúrgico se asociaría a menos complicaciones[6].

- **Autoridad:** Con frecuencia se acude a fuentes especializadas para obtener de ellas el conocimiento que se necesita. Actualmente, esto se ha hecho más común dada la rápida acumulación de información y de avances tecnológicos. En estas circunstancias se tiende a dar credibilidad a quienes tienen mayor autoridad por haber recibido cierto

entrenamiento o por acumular una determinada experiencia. De hecho, es común en este contexto hablar de "expertos" en muchas áreas. Dicha aproximación, sin embargo, comparte con la tradición muchas de sus limitaciones. Por siglos, los expertos sostuvieron que el Sol giraba alrededor de la Tierra, conocimiento del cual luego se demostró que era erróneo.

- Intuición: Esta fuente de conocimiento acepta como criterio de verdad el presentimiento o la corazonada que tiene el individuo de que lo que se cree que es realmente cierto. A pesar de utilizarse a menudo como conocimiento con base en el cual se toman decisiones, tan solo "sentir" que una idea es cierta no constituye un argumento suficiente para aceptar su veracidad.

- Ensayo y error: El método de hacer algo, evaluar sus resultados y repetir el proceso es una aproximación habitual para resolver los problemas, y, por ende, para obtener conocimiento nuevo. Si los efectos son aceptables, puede pasarse a adoptar el método como la forma de solucionar el problema; en caso contrario, se busca otra solución. Si bien tal estrategia, a diferencia de las descritas antes, emplea la verificación para establecer la verdad, su naturaleza poco sistemática puede limitar la validez de sus conclusiones. Adicionalmente, el proceso puede consumir grandes cantidades de tiempo (es necesario probar consecutivamente varias soluciones hasta encontrar la que produce los resultados deseados) o, por el contrario, puede limitar de manera inconveniente la magnitud de la exploración (si la primera alternativa produce resultados "satisfactorios" se suspende el proceso, y se elimina así la posibilidad de probar otras opciones que pueden ser superiores). Por último, podría suceder que se halle una solución aparentemente útil, pero cuyos resultados se deban al azar, y no serían, por lo tanto, repetibles.

- Razonamiento lógico: El razonamiento es una forma de empleo del pensamiento lógico para solucionar un problema que combina la experiencia personal, las facultades intelectuales y sistemas formales de pensamiento. Es un proceso sistemático que se ha empleado a través de la historia para la generación de conocimiento. Se reconocen dos formas de razonamiento lógico: el deductivo y el inductivo.
  - Razonamiento deductivo: Se caracteriza por la aceptación de una proposición general, o *premisa*, y de inferencias subsecuentes que se hacen para casos específicos. El silogismo, introducido por los filósofos de la Grecia antigua, es un ejemplo del método deductivo. Un silogismo contiene una premisa mayor, una premisa menor y una conclusión.

EJEMPLO

Premisa mayor: Todos los seres vivos deben morir.
Premisa menor: El hombre es un ser vivo.
Conclusión: Por lo tanto, el hombre debe morir.

Como puede verse el método se basa en la validez de las premisas, y a menudo se le emplea en ciencias para deducir aplicaciones específicas a partir de verdades más generales. Adicionalmente, este método produce hipótesis que pueden ser verificables mediante la observación empírica. Su principal limitación es que la validez de la conclusión depende por completo de la veracidad de las premisas, que en algunos casos puede ser cuestionable.
Además de lo anterior, este tipo de razonamiento es útil para organizar lo que ya se sabe y para sugerir nuevas relaciones, pero al requerir como punto de partida premisas que ya existen no puede, por definición, crearlas ***de novo***. El razonamiento deductivo es el método empleado por las ciencias formales (es decir, las que se ocupan de conceptos abstractos y de las relaciones que existen entre ellos) para demostrar la validez de sus postulados.

  - Razonamiento inductivo: Este método emplea una lógica inversa a la del método deductivo, pues desarrolla generalizaciones a partir de observaciones específicas. En otras palabras, parte de la experiencia y conduce a conclusiones que, probablemente, son ciertas.

EJEMPLO

La calle se ha mojado cada vez que ha llovido. Por lo tanto, cuando llueva, se mojará la calle.

La estrategia del razonamiento inductivo, que se inició en el siglo XVI con Francis Bacon, fue el origen de la aproximación científica a solucionar problemas. Por ejemplo, la observación de que el cáncer del pulmón es más común en los fumadores que en los no fumadores llevó a postular, a través de la inducción, que el tabaco podía estar asociado al origen de la enfermedad.

El método inductivo también tiene limitaciones; en particular, relacionadas con la validez y la representatividad de las observaciones utilizadas para proponer las generalizaciones. Es imposible estar absolutamente seguro de haber observado todos los posibles casos del evento (por ejemplo, a todos los fumadores y a todos los no fumadores).

A pesar de sus limitaciones, el proceso de razonamiento lógico, en sus formas deductiva e inductiva, es el componente esencial del método científico, que se describe más detalladamente a continuación.

EL MÉTODO CIENTÍFICO

Esta estrategia, que incorpora elementos del razonamiento lógico deductivo e inductivo para el análisis sistemático y controlado de los fenómenos, es la forma más rigurosa de obtener nuevo conocimiento en el campo de la salud.

Se fundamenta en dos presunciones acerca de la realidad: la primera es que la naturaleza es ordenada y regular, y que los eventos son, dentro de ciertos límites, consistentes y predecibles. La segunda es que los eventos no ocurren por azar o por accidente, sino que tienen una o más causas potencialmente identificables.

El método científico tiene varias características:

- En primer término, es ***sistemático*; o sea,** tiene un sentido de orden o de disciplina que busca garantizar un nivel aceptable de reproducibilidad y de validez.
- En segundo término, tiene una ***base empírica*, pues emplea la observación directa para** obtener los datos objetivos necesarios que documentan el conocimiento obtenido.
- En tercer lugar, busca ***controlar* los factores** que no están directamente relacionados con las variables en cuestión, pero pueden influir en ella.
- Por último, emplea el ***examen crítico*, a través** del cual el científico somete sus resultados a la prueba empírica y al escrutinio de otros investigadores, lo que a menudo incluye replicar los hallazgos. Esto le confiere una particularidad muy importante: la de autocorregirse con base en la verificación objetiva, a partir de datos obtenidos en fuentes primarias de información. A menudo, dicha autocorrección se acompaña del empleo de nuevos y mejores métodos para hacer las observaciones, el análisis y las inferencias a partir de los datos.

Esta última característica imprime al saber obtenido a través del método científico una particularidad muy importante: el conocimiento no es infalible ni definitivo, sino que puede variar de manera constante a la luz de nuevos datos y de nuevos hallazgos, que lo modifican, e, incluso, lo hacen obsoleto.

Es decir, a diferencia del conocimiento obtenido por otras de las formas mencionadas antes (la tradición o la autoridad), el logrado a través de la ciencia es refutable.

A pesar de sus ventajas, el método científico no es perfecto. En primer término, hay áreas del conocimiento donde no es aplicable, como la teología o la filosofía. Por otro lado, tiene limitaciones cuando se aplica al estudio de fenómenos muy complejos, tales como el comportamiento humano, en los cuales es mucho más difícil lograr un control suficiente de todas las variables capaces de afectar las observaciones, y que permita descartar la incertidumbre en la interpretación y en la generalización de los datos.

EJEMPLO

Un investigador en el área de ciencias básicas puede controlar muchos aspectos de sus sujetos de investigación. Si trabaja con animales, puede garantizar que sean genéticamente idénticos, que reciban el mismo tipo y las mismas cantidades de alimento, y que se expongan a la misma temperatura, entre otros aspectos. El investigador que trabaja en el campo clínico enfrenta desafíos mayores, pues no puede esperar el mismo grado de control en sus investigaciones, lo cual lo obliga a estar alerta sobre las posibles influencias externas que pueden afectar sus hallazgos.

TIPOS DE INVESTIGACIÓN EN SALUD

La investigación en salud puede clasificarse en diversos tipos, de forma un tanto arbitraria, de acuerdo con el área de interés (y como consecuencia, según los métodos que emplea) y con la aplicabilidad de sus resultados. Es importante señalar que todas estas formas de investigación emplean el método científico y se apoyan en grado diverso en las herramientas desarrolladas por gran variedad de disciplinas como la epidemiología, la estadística, las ciencias sociales, la economía, etc.

- Investigación básica: Este tipo de investigación emplea el método científico para obtener datos empíricos que son de utilidad para desarrollar, refinar o probar teorías. Busca la adquisición de nuevos conocimientos, con la motivación fundamental de hacer nuevos descubrimientos, sin referencia particular a la aplicación práctica de sus resultados. Muy a menudo, este tipo de investigación busca mejorar el conocimiento de los mecanismos celulares, moleculares o genéticos implicados en el mantenimiento de la salud y el desarrollo de la enfermedad. Por supuesto, la investigación básica conduce, a largo plazo, al desarrollo de aplicaciones clínicas o prácticas. Sin embargo, ese no es su objetivo fundamental. Más a menudo que en otras áreas de la investigación en salud, la investigación básica se apoya en laboratorios y en el análisis de especímenes o de muestras obtenidas a partir de seres humanos, animales, microorganismos, plantas, etc.
- Investigación clínica: En este tipo de investigación, el investigador busca, a través de sus observaciones en sujetos humanos intactos, obtener información sobre la frecuencia de las enfermedades, los factores que aumentan el riesgo de padecerlas, los métodos para identificarlas, las intervenciones que pueden ayudar en su prevención o su tratamiento y su pronóstico, entre otros aspectos. A diferencia de lo que ocurre con la investigación básica, en este tipo de investigación se hace más énfasis en la aplicabilidad práctica de los hallazgos para tomar decisiones sobre el cuidado de la salud de individuos.
- Investigación epidemiológica: La investigación epidemiológica busca recolectar información sobre las características y los determinantes de la salud y de la enfermedad en grupos de individuos; es decir, en *poblaciones*. Esto señala su principal diferencia con la investigación clínica, la cual, como ya se discutió, se centra más en sujetos individuales. Sin embargo, al igual que esta, en la investigación epidemiológica se busca obtener información que sea aplicable para tomar decisiones en dichas poblaciones. Es decir, sus resultados también suelen tener aplicabilidad directa.
- Investigación en servicios de salud: Aunque se podría considerar una forma de investigación epidemiológica, la investigación en servicios de salud tiene como objetivo central de su quehacer la forma de proveer atención en salud a las poblaciones y a los individuos, más que la enfermedad misma y sus determinantes. Esta modalidad de la investigación ha crecido de manera considerable, a medida que los sistemas de atención sanitaria aumentan su complejidad y su cobertura, y es particularmente útil para diseñar sus reformas o para evaluar el impacto de la implementación de dichos cambios.

FASES Y PASOS DE LA INVESTIGACIÓN EN SALUD

Este libro presenta a lo largo de sus diversos capítulos una introducción a las principales herramientas metodológicas para conducir investigación en salud empleando el método

científico. La **figura 1.1 muestra un modelo del** proceso de cualquier investigación en el área de la salud, elaborado a partir de lo propuesto por varios autores[2,7,8].

De acuerdo con dicho modelo, una investigación en salud tiene cinco fases: 1) concepción, 2) planificación, 3) implementación, 4) análisis y 5) comunicación.

Estas fases, a su vez, incluyen un número variable de pasos o actividades. Tales pasos pueden, según el diseño o el tipo de estudio específico, variar en su número o en su secuencia. Antes de describirlas, es necesaria una advertencia: la calidad final de la investigación es tan buena como la calidad del más deficiente de todos sus pasos. Cabe aquí la analogía con una cadena, cuya resistencia total es la que tenga el más débil de sus eslabones. Es por ello necesario que el investigador se esfuerce para hacer su tarea de la mejor forma posible en todas las fases que se describen a continuación. Un proyecto perfectamente planificado tendrá resultados de pobre calidad si su ejecución es deficiente.

- Fase 1. Concepción: **La primera fase se centra** en definir y delimitar el tema básico de la investigación; a partir de una idea preliminar, se trabaja para plantear la pregunta y los objetivos del proyecto. Esta fase también incluye la revisión de la literatura, para construir el marco teórico. Todas esas tareas conllevan establecer la justificación de la investigación (llamada también la ***pertinencia*), así como su viabilidad preliminar. Es** importante señalar que esta fase, al igual que cada una de las que le siguen, debe verse como un proceso iterativo, donde a menudo es necesario regresar a los pasos previos a la luz de los desarrollos consecutivos. Por ejemplo, es posible que la revisión de la literatura, o, incluso, el cálculo del tamaño de la muestra (que corresponde a otra fase) hagan necesario volver a considerar la pregunta o los objetivos, por encontrar que investigaciones previas ya han resuelto satisfactoriamente algunos aspectos de los interrogantes originales, o porque el número de sujetos necesarios atenta contra la viabilidad del estudio.
- Fase 2. Planificación: **Esta fase también se** conoce como ***de diseño del estudio*, y** sus pasos tienen como objetivo fundamental definir el plan que se pretende adelantar. De hecho, el resultado del trabajo de las fases 1 y 2 debe ser el protocolo del proyecto, documento fundamental no solo porque sirve de referencia para las fases posteriores, sino porque con él se busca la aprobación de las autoridades de vigilancia (los comités institucionales de ética e investigaciones) y de financiación. En esta fase, el investigador debe establecer, a partir de la pregunta y de los objetivos, la hipótesis del estudio y las variables que le permitirán refutarla o aceptarla, y las cuales, además, tienen que definirse conceptual y operativamente. La pregunta y la hipótesis guían también la selección del diseño o del tipo de estudio más adecuado para responderlas. Los otros pasos de la planificación son definir los métodos para el muestreo y seleccionar a los sujetos de la investigación, así como las estrategias (métodos, observadores y formatos de recolección) para medir las variables en dichos sujetos. Si el diseño seleccionado es un experimento, el investigador debe también establecer la naturaleza de su intervención y los métodos para asignársela a los participantes en el estudio.
- Fase 3. Implementación: **La tercera fase corresponde** a la ejecución del proyecto de acuerdo con el plan establecido durante las fases 1 y 2. Este proceso de recolección de los datos suele ser el que más tiempo consume, y también, cabe reconocerlo, el que menos interés intelectual despierta. Sin embargo, su importancia no es menor que la de las otras fases, pues la calidad de la ejecución es capital para establecer la calidad final del estudio. Muy a menudo (especialmente, en investigaciones complejas), la fase de implementación comienza con un estudio piloto, en el que se someten los planes del estudio a una prueba de terreno. No pocas veces, los hallazgos de esta prueba piloto obligan a hacer modificaciones (idealmente, de carácter menor) en los procedimientos, lo cual refuerza la idea de que la secuencia descrita en la figura es iterativa y flexible.
- Fase 4. Análisis: **En esta fase, el investigador** procede a emplear las estrategias más

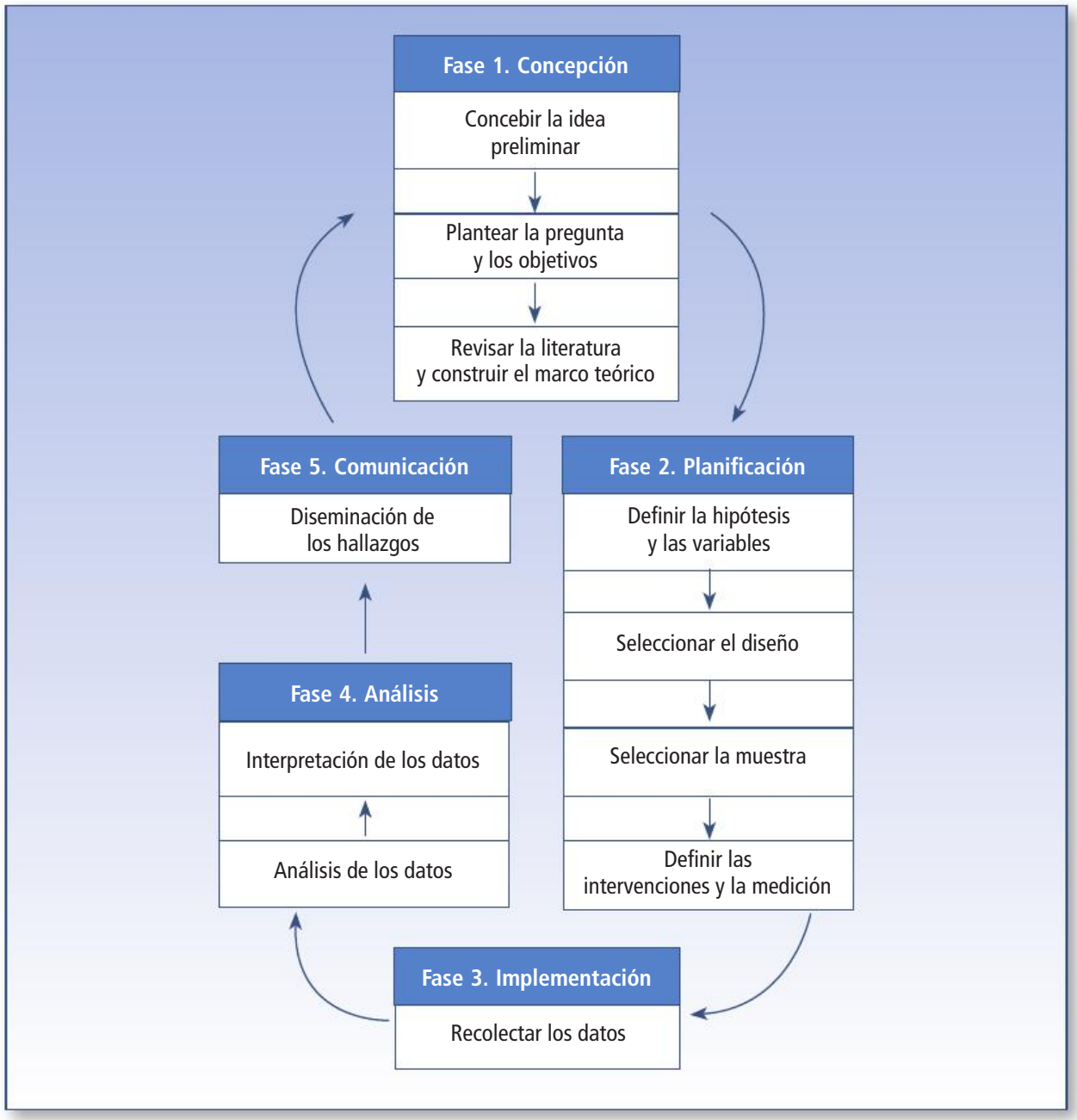


Figura 1.1 Fases y pasos de una investigación.

adecuadas para depurar y resumir los datos obtenidos a lo largo de la fase precedente. Las técnicas más apropiadas dependen de la naturaleza de la pregunta y de la información recolectada. En muchos casos, durante esta fase se acude a la estadística como herramienta clave del proceso. El hecho de que solo hasta este punto se hable del análisis no quiere decir que la propuesta del estudio desarrollada al final de las fases 1 y 2 no deba incluir el plan de análisis. Finalmente, una vez se haya analizado la información, el investigador debe interpretar los hallazgos a la luz de la pregunta y de los objetivos originales, así como de lo encontrado previamente por otros proyectos similares. Esto último implica volver a la revisión de la literatura de la fase 1, lo que a menudo conlleva su actualización; en especial, si la fase de implementación tomó mucho tiempo.

- Fase 5. Comunicación: La última fase del proceso corresponde a la diseminación de los

hallazgos. La forma más común para ello es elaborar documentos que se someten a publicación en forma de artículos, de capítulos o de libros, así como presentar los hallazgos en reuniones científicas. Si bien existe la convicción de que la comunicación de sus hallazgos es una "obligación" del investigador[9], la evidencia acumulada indica que, a menudo, esta fase se cumple de manera incompleta y tardía, lo cual implica que los hallazgos de la investigación en salud toman un tiempo inaceptablemente prolongado para su implementación en la práctica[10].

Por último, es necesario señalar que ninguna investigación es el final del camino. Los resultados de un estudio generan nuevas preguntas, que requieren, a su vez, investigaciones adicionales para responderlas. Por ello, la **figura 1.1** presenta las fases de la investigación en la forma de un ciclo.

La investigación en salud busca mejorar, a través de la generación de conocimientos de calidad impecable, la salud y la calidad de vida de los seres humanos. Su fin no es producir conocimientos generalizables, sino que estos sean aplicables a la atención en salud; idealmente, en poblaciones amplias.

La calidad de la información obtenida depende, estrictamente, de la calidad de cada uno de los pasos que deben seguirse en la investigación, por lo cual el investigador debe asegurarse de que cuida con la mayor atención cada uno de dichos pasos.

Referencias

1. Commission on Health Research for Development. Health research: essential link to equity in development. Oxford: Oxford University Press; 1990.
2. Portney LG, Watkins MP. A concept of clinical research. En: Portney LG, Watkins MP, editores. Foundations of clinical research. Applications to practice. 3rd. ed. Upper Saddle River: Prentice Hall; 2008. p. 3-20.
3. Herrera AN. Sánchez R, Pardo R. El método científico. En: Ardila E, Sánchez R, Echeverri J, editores. Estrategias de investigación en medicina clínica. Bogotá: El Manual Moderno; 2001. p. 1-7.
4. Ebrahim GJ, Sullivan KR. The scientific method. En: Ebrahim GJ, Sullivan KR, editores. Mother and child health research methods. Trowbridge: Redwood Books; 1995. p. 1-13.
5. Dohoo I, Martin W, Stryhn H. Introduction and causal concepts. En: Dohoo I, Martin W, Stryhn H. Methods in epidemiologic research. Charlottetown: VER Inc.; 2012. p. 1-34
6. Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. BMJ. 2012;344:e2156.
7. Hulley SB, Newman TB, Cummings SR. Getting started: the anatomy and physiology of clinical research. En: Hulley SB, Cummings SR, Browner WS, et al., editores. Designing clinical research. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 2-13.
8. Hernández R, Fernández C, Baptista P. Prólogo. En: Hernández R, Fernández C, Baptista P. Metodología de la investigación. 3ra. ed. México: McGraw Hill; 2003. p. XXXI-VII.
9. Chalmers I. Underreporting research is scientific misconduct. JAMA. 1990;263:1405-8.
10. Lenfant C. Clinical research to clinical practice. Lost in translation? N Engl J Med. 2003;349:868-74.



CAPÍTULO 2

La pregunta de investigación

JUAN GABRIEL RUIZ PELÁEZ

INTRODUCCIÓN

El objetivo de la investigación científica es generar conocimiento científico; o sea, responder preguntas utilizando el método científico. En el presente capítulo se busca establecer una definición práctica del conocimiento científico y describir y analizar la importancia y las características de la formulación de preguntas de investigación como un paso primario para hacer investigación científica. También se explora el origen de las preguntas, así como sus atributos y su clasificación, y se ofrecen al lector algunas estrategias específicas para ayudarle en la formulación de preguntas e hipótesis científicas que sean apropiadas, eficientes y útiles.

EL CONOCIMIENTO CIENTÍFICO

El debate planteado entre dos corrientes filosóficas, el ***monismo*** (solo existe el mundo físico y todos los fenómenos son, en últimas, términos físicos) y el ***dualismo*** (existen el mundo físico y un mundo inmaterial —el de las ideas y la emociones—), parece resolverse a favor de la segunda cuando se plantea en términos de información. Por ejemplo, los seres vivos son la síntesis de la dualidad entre forma y función (***fenotipo***) y de la información sobre forma y función (***genotipo***). Reconocemos a un individuo como parte de una especie porque reúne una serie de características que son típicas de dicha especie. La esencia de tales características está contenida en la información genética. El viejo problema de qué fue primero, si el huevo o la gallina, puede reformularse así: los genes son la manera como los seres vivos (la forma, la expresión concreta) se perpetúan, o bien, los seres vivientes son la manera como los genes (la idea, la información "abstracta") se mantienen y evolucionan.

Karl Popper, uno de los más influyentes filósofos de la ciencia en el siglo XX, va un paso más allá del monismo y el dualismo. Popper propone pensar en el conocimiento adquirido por la consciencia como el producto de la interacción de tres "mundos": 1) el ***mundo uno***, o ***mundo físico*** (el de los objetos, el universo tangible), donde hay entes y eventos que pueden ser percibidos por un observador; 2) el ***mundo dos***, donde cada sujeto se hace una representación mental de lo que ocurre en el mundo físico, a través de sus percepciones y sus interpretaciones, y donde, además, residen las emociones, los recuerdos, las experiencias, etc. (lo ***subjetivo***), y, finalmente, 3) el ***mundo tres:***

el mundo de las ideas y los conceptos, y en el cual diferentes observadores del mundo físico, que tienen su propia consciencia y su mundo subjetivo (mundo 2), encuentran o construyen el conocimiento racional. Gracias a que los seres humanos tienen lenguaje, el cual es un sistema complejo de símbolos y signos, diferentes observadores pueden compartir y comunicar sus experiencias, sus percepciones y sus ideas, y construir (o descubrir) leyes, teorías, obras de arte, sistemas lógicos y matemáticos, etc. Los "entes" que habitan dicho tercer mundo tienen existencia objetiva, y diferentes observadores pueden apreciarlos, juzgarlos y comprenderlos. Ese descubrimiento o creación de los habitantes del mundo tres es lo que se consolida como conocimiento o como cultura.

Independientemente de si existen una o varias realidades objetivas que se manifiestan, en términos prácticos, nuestras conciencias solo interactúan con la información proveniente de la realidad (mundos uno y tres de Popper); y con base en esa información, nos hacemos una representación (idea) de esa realidad "externa". Todo lo que existe es afectado por información proveniente de su entorno, y reacciona y se adapta a dicha información. Los cuerpos astronómicos intercambian gravitones, y en respuesta a dichas interacciones (a la información intercambiada) describen sus órbitas y se desplazan por el espacio. La emisión y la recepción de información (el ejercicio de mutuas influencias y sus respuestas adaptativas) parece ser un proceso automático y determinístico. Reaccionar a la información no implica que el ente tenga conciencia de dicha realidad que lo afecta; de hecho, estas reacciones no necesariamente son libres ni deliberadas.

Cuando trasladamos el problema a las interacciones de los seres vivos con su ambiente (que incluye otros seres vivos), su descripción se torna más compleja. Sin embargo, solo reconocemos el procesamiento consciente de información, que permite comprender reglas generales y formular predicciones, en el ámbito de los seres humanos. Reconocemos la conciencia (percepción del entorno bien sea externo u objetivo, interno o subjetivo y cognitivo —mundo tres—, junto con la capacidad para recordar el pasado, para anticipar el futuro y para reconocerse a sí mismo) como una característica humana, y reservamos el término "conocimiento" al producto del procesamiento consciente de información, además de la clasificación y el archivo de información o el desarrollo de conductas adaptativas, que implica la abstracción de conceptos (resumen del recuerdo o de los eventos pasados), su aplicación a situaciones presentes o futuras, y la toma de decisiones, basadas no simplemente en la información explícita, sino en procesos mucho más complejos, y que excede la simple capacidad para sortear situaciones (solución de problemas, que es una definición popular de inteligencia).

DEFINICIONES

El término "conocimiento" se utiliza para expresar diferentes conceptos. El diccionario de la Real Academia de la Lengua define ***conocimiento*** **como "entendimiento, inteligencia,** razón natural". ***El conocimiento*** **(con artículo** definido) se refiere al concepto de ciencia o saber. En el sentido utilizado en este capítulo, conocimiento significa tanto la información adquirida por la conciencia (descripción) como la explicación que la conciencia se ofrece a sí misma de los hechos o de los fenómenos descritos por dicha información. El conocimiento es, para la conciencia, la sensación de que se tiene información sobre objetos y fenómenos (datos, observaciones, conceptos) y se tiene una explicación de cómo y por qué interactúan los objetos y se producen los fenómenos. Si los datos son correctos y la explicación es apropiada, el conocimiento debería permitir que se prediga o se anticipe, y que, por tanto, se controle, en alguna medida, el devenir de las cosas.

Qué tan verdadero es el conocimiento (es decir cuál es la veracidad de las descripciones y de las explicaciones ofrecidas, y, por ende, su capacidad de ajuste a los hechos —su concordancia con la "verdad"—) es uno de los problemas centrales de la ciencia. La veracidad y la utilidad del conocimiento van a depender de la calidad de la información que se tiene y de la validez de los raciocinios que analizan y sintetizan dicha información.

Los seres humanos solemos guiarnos por un conjunto implícito y explícito de creencias, valores y conocimientos. Los epistemólogos identifican dos grandes tipos de conocimiento cuya validez es diferente: el conocimiento "común", o "vulgar", y el conocimiento científico. Estas diferentes formas de adquirir y validar

conocimiento van a influir en nuestras creencias y nuestros valores, y, a su vez, estas creencias y valores pueden (y suelen) moldear qué aprendemos y cómo conocemos.

El conocimiento común se adquiere, en gran medida, a través de la experiencia cotidiana. La acumulación de experiencias similares lleva a hacer generalizaciones que constituyen conocimiento común. Por ejemplo, la experiencia indica que el agua se mantiene líquida a temperatura ambiente, mientras que se evapora cuando se la calienta, lo cual lleva a establecer la generalización de que el agua ***siempre*** se mantiene líquida a temperatura ambiente, y ***solo*** se evapora al calentarla.

Pero algunas circunstancias inusuales pueden contradecir estas generalizaciones. Por ejemplo, el agua puede evaporarse a temperatura ambiental si la presión atmosférica disminuye mucho; incluso, el hielo puede pasar directamente a la fase gaseosa en el vacío (sublimación). Sin embargo, en la vida cotidiana es improbable que se presente una situación donde el hielo se sublime, y debido a ello se tiende a aceptar como universalmente válido el ***postulado*** de que el agua solo se evapora al calentarla. Dicho postulado no explica el comportamiento del agua: simplemente, lo describe y predice lo que ocurrirá con el agua, de manera satisfactoria y en circunstancias usuales. El estudio científico de los estados de la materia permite explicar por qué el agua se mantiene en estado líquido si se cumplen ciertos requisitos de *temperatura* y de *presión*.

El conocimiento científico permite predecir circunstancias inusuales en las que el agua se comportará de manera diferente de como lo postula el conocimiento común. Ver el agua hirviendo a temperatura ambiente es un evento bizarro que contradice la experiencia cotidiana; va contra el ***sentido común,*** pero la razón puede aceptarlo como valedero.



El conocimiento común es relativamente ingenuo, y se sustenta en el "sentido común".

Esta fundamentación se basa en tres fuentes:

1. La percepción directa mediante los órganos de los sentidos de los fenómenos. Dicha percepción es tanto ***concurrente*** (presente) como ***pasada*** (experiencia propia).
2. El raciocinio propio (argumentación usualmente inductiva), que genera explicaciones de "sentido común".
3. La experiencia de otros.



El conocimiento común tiende a ser dogmático y puede apoyarse en interpretaciones no racionales (pensamiento *mágico*).

Los atributos negativos usualmente identificables en el conocimiento vulgar, o común, incluyen:

a. Es ***superficial***, pues aunque parte de percepciones de los hechos, no explora sus relaciones profundas, sino que se queda con las aparentes, e, incluso, puede interpretarlas de manera no racional (por ejemplo, pensamiento mágico: las cosas ocurren por mala suerte, por maleficios, etc.).
b. Es ***subjetivo***, y tanto la integración de los datos como la explicación ofrecida dependen de características y particularidades del individuo que lo genera, y es muy susceptible a colorearse o distorsionarse por emociones, experiencias previas negativas o positivas, ideologías y prejuicios.
c. Tiende a ser ***dogmático***, a fundamentar sus explicaciones en declaraciones aceptadas en sí mismas, independientemente de su veracidad y su validez.
d. Es ***estático;*** o sea, no evoluciona con facilidad, y los nuevos hechos no necesariamente hacen cambiar las explicaciones ya aceptadas.
e. Es ***marcadamente particular.*** No se suele buscar que la explicación ofrecida sea consistente con fenómenos análogos.
f. Es ***poco sistemático,*** ya que los datos y las explicaciones ofrecidas no necesariamente se clasifican de una manera sistemática y consistente, ni obedeciendo a un orden lógico
g. Es ***inexacto,*** tanto en la reproducibilidad de la información como en la precisión y la claridad de los términos usados para expresar conceptos y explicaciones.
h. Usualmente, ***no es acumulativo;*** las explicaciones no se agregan en un cuerpo creciente, sino que se apilan de manera caótica. De hecho, se puede dar el caso de que pueden coexistir explicaciones contradictorias para eventos análogos y pueden ser utilizadas al tiempo por el mismo individuo.

Por el contrario, el conocimiento científico se fundamenta en hechos y datos observados cuidadosamente y de forma estandarizada, y busca ofrecer explicaciones científicas.

El objeto del conocimiento científico (las preguntas de investigación) puede coincidir con el del conocimiento vulgar, aunque lo contrario no siempre ocurre. Los atributos deseables del conocimiento científico se contraponen a los del vulgar:

a. Es *racional:* explora más allá de los hechos y de sus relaciones aparentes, e intenta establecer, de manera sistemática y utilizando los principios del raciocinio lógico, las razones que explican la información aparente.
b. Es *objetivo:* evalúa hechos y datos percibidos como externos al observador, los analiza y los contrasta.
c. Es *flexible:* está dispuesto a aceptar explicaciones lógicas y sustentadas en datos, aunque contradigan creencias o expectativas. Suele mirar las explicaciones ofrecidas como tentativas y transitorias, y está dispuesto a corregirlas o reemplazarlas.
d. Es *evolutivo y acumulable:* el conocimiento adquirido se articula de manera lógica con el conocimiento precedente, y facilita la adquisición de conocimiento futuro.
e. Abstrae y puede generalizarse para extender las explicaciones a datos y a hechos diferentes o aún no observados.
f. Es *sistemático* y usa un lenguaje riguroso y preciso.

Además de los atributos enumerados, las diferencias entre el conocimiento vulgar y el científico deben explorarse en: a) el ámbito del objeto que se quiere conocer y b) el tipo de explicación ofrecida a la naturaleza y el comportamiento del objeto.

En lo que respecta a la naturaleza misma del objeto del conocimiento, este puede ser de cualquier índole en el conocimiento vulgar, mientras que en el conocimiento científico debe tratarse de entes o fenómenos susceptibles de ser percibidos, observados y medidos cuantitativamente o cualitativamente. Es decir, los tópicos de adquisición de conocimiento, los problemas que surgen en dichos tópicos y las eventuales preguntas en torno a los problemas planteados deben relacionarse con eventos perceptibles y mensurables objetivamente.

Hay otros tópicos susceptibles de ser evaluados mediante el raciocinio lógico y crítico (distinto del conocimiento vulgar), pero que no se refieren a eventos y entes de la realidad externa u objetiva, y, por tal motivo, se salen del dominio de las ciencias y entran en el de otras disciplinas racionales, como la filosofía. Típicamente, el método científico no es apropiado para responder preguntas acerca de principios y valores, de ética y moral, o de lo sobrenatural y trascendente. Esto de ninguna manera implica que el científico, como ser humano, se margine de la axiología, la ética o la religión.

Respecto al tipo de explicación ofrecida, no necesariamente hay una diferencia tajante y cualitativa entre la explicación cándida basada en el sentido común y la llamada explicación científica. Las diferencias son más de magnitud que de atributo. Por ejemplo, las explicaciones no científicas tienden a ser *dogmáticas;* es decir, aceptadas como verdaderas y no susceptibles de ser corregidas o mejoradas. Por el contrario, se espera que las buenas explicaciones científicas se alejen de esta rigidez; sin embargo, hay muchas instancias donde en el saber común las explicaciones se expresan como tentativas y provisionales (en espera de corrección o de verificación), mientras que, lamentablemente, a veces, las supuestas explicaciones científicas se plantean como absolutas e incontestables; o sea, dogmáticas.

Uno de los atributos más importantes y específicos de la explicación científica es el hecho de que se deriva de un proceso de raciocinio lógico y racional. El proceso de pensamiento crítico se fundamenta en dos grandes pilares: a) un uso apropiado del lenguaje y b) la disposición de una línea de argumentación lógica que sea válida. En los siguientes apartes se describen brevemente ambos aspectos.

Uso del lenguaje en los procesos de explicación científica

Los expertos en argumentación lógica reconocen al menos tres usos del lenguaje:

a. El *uso expresivo* del lenguaje, que busca ventilar un sentimiento o emoción, o, incluso, evocarlo en otros. La afirmación "Los lunes son días agobiantes" puede originarse en hechos: el contraste entre el descanso

dominical y el reinicio de las actividades diferentes del ocio, la eventual acumulación de trabajo no resuelto durante el fin de semana, etc.; pero el mensaje que se comunica no se refiere a esos hechos, sino a un sentimiento de agobio o de angustia que puede o no estar relacionado con ellos.

b. **El uso *directivo*, o *prescriptivo*, del lenguaje:** las palabras buscan generar o impedir una acción determinada por parte de un agente (por lo general, humano): "Siga al consultorio", "Debe dejar de fumar", sin comunicar un contenido. Por ejemplo, no se discute la razón para entrar al consultorio; tampoco se comunican los argumentos para dejar de fumar, que, de hecho, pueden ser de cualquier índole: porque es perjudicial para la salud, o, simplemente, porque está prohibido.

c. **Uso *informativo* del lenguaje:** trata de comunicar un contenido; generalmente, hechos, datos o conceptos. Este uso del lenguaje tiene sentido cuando lo que se intenta informar se presume verdadero.

En la argumentación lógica que sustenta la formulación de explicaciones científicas (y que está en la base de la construcción del conocimiento o saber científico) es muy importante utilizar un lenguaje preciso, y con intención informativa.

Es necesario tener presente que muchas palabras, además de su significado literal, pueden también, en circunstancias tanto generales como específicas, tener un significado emocional particular. El uso de palabras que puedan tener una carga emocional específica puede distorsionar su significado literal, y generar discrepancias aparentes. Por ejemplo, al describir los procesos de argumentación científica como "racionales", algún lector puede interpretarlo como "adscritos a la escuela racionalista", y si su posición ideológica es contraria al racionalismo, es muy posible que emocionalmente se sienta en desacuerdo, aunque esté de acuerdo con los hechos expuestos. Es posible que si se emplea la palabra "lógicos", aun cuando el significado literal no es equivalente al de "racionales", al evitar la segunda palabra, que tiene un sentido emocional negativo, se logre comunicar mejor el concepto que se quiere transmitir.

Para un uso efectivo del lenguaje en el conocimiento científico (sobre todo, en la formulación de preguntas e hipótesis científicas y de sus predicciones), no basta con emplear términos claramente definidos, sino que también se debe tratar de escoger palabras y frases con la menor carga emocional posible.

El resultado, aunque algo árido e insípido, es mucho más apropiado para una adecuada comunicación de los procesos y de los contenidos del conocimiento científico.

En el capítulo acerca de cómo escribir y publicar reportes de investigación se retoman y se amplían estas consideraciones sobre usos del lenguaje y sobre la escogencia del tipo de lenguaje más apropiado para la comunicación científica.

Argumentación lógica

Para ofrecer explicaciones a hechos observados o percibidos, es necesario establecer procesos de razonamiento que, eventualmente, lleven hacia algún tipo de conclusión.

La explicación científica utiliza argumentos lógicos. Los ***argumentos lógicos*** se componen de dos o más proposiciones relacionadas entre sí. Una ***proposición*** es un enunciado al que se busca afirmar o contradecir (establecer si es verdadero o no).

Las proposiciones lógicas solo pueden adoptar uno de dos valores: ser ***verdaderas*** o ***falsas,*** independientemente de si en un momento dado existe información para establecer lo uno o lo otro.

En un argumento lógico hay una o más proposiciones que, supuestamente, proveen soporte a otra proposición. La primera o las primeras constituyen las ***premisas,*** y la segunda es la ***conclusión.*** La manera como la conclusión se desprende de la relación entre las premisas, la forma como se transita de premisa a conclusión, la conexión lógica entre dichas proposiciones, es la ***inferencia.***

Por ejemplo, las frases "en verano hace calor", "el clima es difícil de predecir" y "en los

desiertos la precipitación es escasa" son proposiciones que pueden ser falsas o verdaderas, y que aun cuando se refieren al tópico del clima, no constituyen un argumento, pues no tienen relación de premisa y conclusión entre ellas. Por el contrario, las proposiciones "la encefalitis rábica es invariablemente letal", "el paciente fue mordido por un perro rabioso" "el paciente sobrevivió" y "el paciente no desarrolló encefalitis rábica" conforman un ***argumento,*** donde las tres primeras proposiciones son las premisas, y la última es la conclusión que se infiere de ellas.

Es necesario establecer la coherencia y la confiabilidad de los procesos de inferencia, de los patrones que llevan de las premisas hacia las conclusiones en los argumentos lógicos. Es necesario comprender cuáles de dichos patrones son coherentes y confiables (inferencia adecuada) y cuáles no. En principio, hay dos grandes tipos de inferencia, que tienen estructuras y normas propias y distintivas: 1) la ***inferencia*** *deductiva* y 2) la ***inferencia inductiva.***

Habitualmente, se llama ***deductivo*** a lo que parte de principios generales para arribar a conclusiones particulares; esto, si bien es correcto, no necesariamente es lo que define a la inferencia deductiva. La inferencia deductiva ocurre cuando un argumento dice poder garantizar la veracidad de su conclusión basado en la veracidad de sus premisas. La inferencia deductiva es exitosa solo si sus premisas proveen un apoyo tan absoluto y completo que es por completo inapropiado e inconsistente (es decir, "ilógico") suponer que las premisas son verdaderas, pero la conclusión no. Un ***argumento deductivo*** solo puede ser completamente exitoso o no serlo; no hay valores intermedios. Las ***estructuras*** (los patrones de relación entre premisas en algunos tipos de argumentos deductivos) son incontestables y perfectas, y si sus premisas son, de hecho, verdaderas, la conclusión solo puede ser verdadera, independientemente de cualquier otra circunstancia. Todos los otros tipos de argumentos deductivos carecen de utilidad, ya que, independientemente de la veracidad de sus premisas, las conclusiones pueden ser falsas, y la adquisición de información adicional no ayuda a cambiar este hecho.

Por su parte, el ***razonamiento inductivo*** es entendido como el que parte de observaciones particulares e intenta inferir conclusiones generales. El paso de lo particular a lo general no permite aseverar de manera absoluta que si las premisas son verdaderas la conclusión también lo será, pues la conclusión hace afirmaciones más generales que las contenidas en las premisas. Por ejemplo, observando instancias repetidas de la mejoría de pacientes con neumonía bacteriana después de recibir antibióticos (premisas particulares), se puede inferir la conclusión de que los antibióticos siempre mejorarán a los pacientes con neumonía. Sin embargo, es posible concebir situaciones donde las premisas sean ciertas (en todos los casos que se han observado, los antibióticos se asociaron a curación de la neumonía), y, sin embargo, la conclusión no se cumpla; es decir, en instancias futuras o no detectadas podría haber casos de neumonías que no mejoren. Este problema ha sido conocido como la ***falacia de la inducción.***



Con base en la inferencia inductiva, por ende, un argumento no puede establecer la certeza de su conclusión; sin embargo, puede afirmar que la conclusión tiene probabilidades de ser correcta.

A diferencia de la deducción, en la inducción hay grados variables de verosimilitud de las conclusiones, y la adquisición de nueva información (premisas) puede afectar el grado de confianza que se tenga en la verosimilitud de la conclusión.

Para formular preguntas de investigación (en particular, la generación de hipótesis adecuadas y verificables), se requiere el uso de procesos de inferencia o de predicción tanto inductiva como deductiva. La deducción de predicciones (conclusiones esperadas) es un paso muy importante en la investigación científica. Por tal motivo, se examinarán aquí, en algún detalle, los conceptos de ***validez*** y ***veracidad*** de los argumentos deductivos, así como las falacias derivadas de imperfecciones de diferente naturaleza en estos procesos.

Veracidad y validez de los argumentos deductivos



Un argumento deductivo es válido si la inferencia que deriva la conclusión a partir de las premisas es perfecta.

Esta afirmación puede frasearse de dos maneras lógicas (que se deducen de ella): a) en un argumento válido, si las premisas son verdaderas, la conclusión solo puede ser verdadera; y b) es imposible que la conclusión sea falsa mientras sus premisas sean verdaderas.

Cualquier argumento deductivo que no sea válido es inválido: es posible que la conclusión sea falsa aun cuando las premisas sean verdaderas; es decir, aun en el evento de que las premisas sean verdaderas, la conclusión puede ser verdadera o falsa. Un argumento inválido no permite adquirir conocimiento.

La validez del argumento es independiente de la verdad de las premisas. Para que la conclusión sea verdadera, deben cumplirse ambas condiciones; es decir, que la inferencia sea valedera y que las premisas sean ciertas.

De las ocho posibles combinaciones entre veracidad y validez, solo en una puede descartarse la falsedad de la conclusión, tal como se ilustra en la **tabla 2.1**.

Hay una serie de errores en la argumentación deductiva que llevan a inferencias inválidas, donde el argumento no consigue proveer un soporte lógico adecuado para sustentar la verdad de su conclusión; sin embargo, parecería ser convincente o persuasivo de alguna otra manera. Estos errores son conocidos como falacias, y los lógicos las han tipificado en dos grandes grupos:

a. Las *falacias formales,* causadas por errores estructurales en la lógica deductiva.
b. Las *falacias informales,* que derivan su aparente plausibilidad de un uso inadecuado del lenguaje ordinario, y a través de las cuales se intenta persuadir de la verdad de la conclusión usando argumentaciones no racionales. Las falacias informales se clasifican, a su vez, en tres grupos: 1) presunción, 2) relevancia y 3) ambigüedad.

Tabla 2.1 Relaciones entre la veracidad de las premisas, la validez de las inferencias y la veracidad de las conclusiones en los argumentos deductivos

Premisas	Inferencia	Conclusión
Verdadera	Válida	Verdadera

	Inválida	Verdadera
		Falsa
Falsa	Válida	Verdadera
	Inválida	Verdadera
		Falsa

El conocimiento común suele verse comprometido por el uso de argumentos en apariencia lógicos, pero que, en realidad, son falacias informales. Por su parte, el conocimiento científico (en particular, la formulación de explicaciones científicas) descansa, sin duda, en la argumentación lógica, y, por ello, el investigador debe utilizar correctamente los principios elementales de la lógica para establecer líneas de argumentación apropiadas, evaluar críticamente la validez y la veracidad de las explicaciones (hipótesis y teorías) y reconocer las falacias, tanto formales como informales, y otras fuentes de error en el raciocinio crítico.

LA PREGUNTA DE INVESTIGACIÓN

Formular una pregunta de investigación es un paso indispensable en el proceso de diseño y desarrollo de una investigación. La falta de preguntas claras y bien formuladas suele implicar la conducción de ejercicios investigativos vagos, confusos y cuyos resultados son difíciles de interpretar. Sin embargo, no basta con formular una pregunta científicamente relevante y valedera, y susceptible de ser contestada mediante la observación de fenómenos naturales. Además de ello, es necesario poner en funcionamiento dicha pregunta, de una forma sistemática, hasta construir un proceso válido para aproximarse a su contestación. Usualmente, la pregunta de investigación se refina hasta formular objetivos (metas) específicos. De la claridad de la formulación de los objetivos dependerá, en gran medida, la capacidad de los desarrollos ulteriores del proceso investigativo.

Tópicos de investigación *vs.* problemas de investigación

El hecho de hacer una investigación puede originarse en la simple curiosidad desprovista de algún componente utilitario, o, como contraparte, en la necesidad de solucionar problemas. Independientemente de qué motivación lo impulse, el investigador eficaz necesita un grado considerable de foco tanto de contenido, como disciplinar y metodológico. Formular preguntas

adecuadas, pertinentes y susceptibles de ser contestadas requiere conocimiento previo sobre el tópico general, y conocimientos y destrezas específicos en al menos un área disciplinar.

Hay al menos dos enfoques para generar preguntas de investigación: 1) el ***centrado en el tópico*** o 2) el ***centrado en el problema.***

En el primero de los enfoques mencionados, el investigador, apoyado en una ciencia o en una disciplina, e interesado en un tópico particular y familiarizado con él, busca baches de conocimiento o situaciones y eventos específicos a partir de los cuales generará sus preguntas de investigación. Por ejemplo, el profesional con entrenamiento en genética humana e interés y profundización previa en el tópico de los errores congénitos del metabolismo de los carbohidratos es muy probable que genere preguntas de investigación originadas en su conocimiento del tópico que le permite identificar vacíos de información, o conocimiento o problemas específicos que atraen su atención y la de sus colegas disciplinares. Este enfoque tiende a producir preguntas muy enfocadas, con gran relevancia intradisciplinar, pero también, posiblemente, de menos relevancia inmediata para el ejercicio profesional. Las preguntas originadas en tópicos generan respuestas enfocadas en el cuerpo de conocimientos teóricos de una ciencia; es decir, tienden a alinearse con las llamadas "ciencias puras". En el caso de las ciencias de la salud, las preguntas generadas en este primer enfoque derivan más hacia el área de las ciencias básicas, y buscan describir y explicar, más que solucionar, problemas (***enfoque explicativo***).

En el segundo enfoque, el investigador tiene una actividad en un área profesional específica o que está relacionada con la problemática profesional del área, y su actividad profesional lo enfrenta a vacíos de conocimiento o a la necesidad de comprender, explicar y modificar situaciones o solucionar problemas. Los médicos e investigadores clínicos y epidemiológicos suelen encajar mejor en este enfoque. Las preguntas no se originan en un tópico o una disciplina, sino en torno a una actividad profesional que tropieza con un problema y debe solucionarse.

El clínico que trata infecciones urinarias se ve ante el problema de que, pese a enfrentarse a pacientes con riesgos en apariencia similares, con manejo terapéutico aparentemente comparable, en algunos casos presentan cicatrices renales y pérdida acumulativa de la función renal, y en otros, no. El problema tiene múltiples facetas, y su abordaje difícilmente se puede hacer con una perspectiva unidimensional.

El análisis del problema lleva a identificar una serie de aspectos, donde cada uno de ellos es una posible fuente de preguntas específicas de investigación que deben abordarse con conocimientos y metodologías provenientes de más de una disciplina. La variación observada en la evolución de la infección urinaria, por ejemplo, ¿tiene que ver con aspectos microbiológicos, genéticos, metabólicos o comportamentales? Las preguntas originadas en este enfoque tienden a ser más pragmáticas que explicativas, aunque ambos tipos de aproximación se pueden dar en enfoques centrados en tópicos o centrados en problemas.

El presente libro está orientado, fundamentalmente, hacia la investigación clínica y epidemiológica, y, por ende, en él se explorarán, de preferencia, el origen y la formulación de preguntas de investigación en el enfoque centrado *en problemas.*

Adicionalmente, la motivación de los investigadores puede provenir del impulso autónomo de conocer, o bien, de la necesidad de investigar impuesta desde afuera. Algunos de los lectores (en particular, los estudiantes de pregrado y de posgrado que tienen la necesidad de hacer investigaciones para trabajos de grado o como parte de sus tesis) estarán en esa segunda categoría al menos durante sus primeras aproximaciones al ejercicio investigativo. Para ellos puede resultar intuitivamente más adecuado acercarse a formular preguntas de investigación centrándose en un tópico; por ejemplo: "quiero investigar en dermatitis atópica". Sin embargo, suele ser más productivo y eficiente centrar el foco en la identificación de problemas: "¿Qué hacer para que los niños con antecedentes familiares de atopía no desarrollen dermatitis atópica?".



Las respuestas a preguntas originadas en tópicos tienden a ser ***explicativas;*** las originadas en problemas tienen a ser ***pragmáticas.***

Fuentes de preguntas de investigación centradas en problemas

En la atención en salud hay una serie ordenada de procesos que implican tomar decisiones.

Cada toma de decisiones requiere información adecuada y un proceso de evaluación crítica de opciones. Cada uno de estos puntos relativos a tomar decisiones en el proceso de atención es una fuente potencial de preguntas investigativas.

De una manera simplificada, se puede definir la actividad profesional clínica como la toma y la implementación de decisiones relativas al mantenimiento de la salud y la promoción del desarrollo de las potencialidades de la gente (promoción de la salud), la prevención o el retardo de la aparición de enfermedades o de daños (prevención primaria), la prevención o la atenuación de la aparición de complicaciones (prevención secundaria) o de secuelas (prevención terciaria) y, de manera más global, la restauración de la salud (tratamiento) y la función (rehabilitación) y, en general, la prolongación de la vida con la adecuada calidad. En todos estos puntos, las intervenciones producto de la toma de decisiones recaen en dos grandes áreas: modificar condiciones o eventos (intervenciones preventivas, terapéuticas o de rehabilitación) o recoger información adicional (interrogatorio, examen físico, tamizado, pruebas diagnósticas clínicas, paraclínicas e imaginológicas, etc.).

Puede formularse una gran cantidad de preguntas relevantes sobre intervención (eficacia, efectividad, eficiencia, costo), sobre información adicional (habilidad discriminante de pruebas diagnósticas, costo-efectividad, utilidad, aceptabilidad, etc.) o sobre condiciones basales (demografía, distribución de factores de riesgo y pronóstico, acceso a servicios de salud, comportamiento, etc.).

En general, las preguntas de investigación centradas en los problemas de atención reflejan los tópicos de interés en la relación clínica, como puede verse en la **tabla 2.2**.

La necesidad de contestar preguntas en el curso de la práctica suele aparecer cuando el investigador percibe alguna variabilidad inesperada en la frecuencia de una condición, en los resultados de maniobras terapéuticas o diagnósticas en medio del curso clínico o en el pronóstico de una condición. En otras ocasiones, la investigación se origina en la necesidad de contestar interrogantes sobre la enfermedad o su tratamiento, que surgen de los pacientes o de las experiencias de estudiantes y colegas. Otra circunstancia asociada a menudo a la generación de preguntas de investigación es que se introduzcan nuevos métodos diagnósticos, modalidades terapéuticas o tecnologías en salud.

Clases de preguntas de investigación

La investigación científica puede definirse como el uso del método científico para responder preguntas, y así generar conocimiento científico. En la discusión previa sobre el conocimiento científico se identificó que este tiene al menos dos grandes componentes: *información* (datos, descripciones o "mapas") e *inferencias* (predicciones y explicaciones).

En este punto, es necesario llamar la atención sobre los dos niveles de las inferencias: 1) ***predecir*** y 2) ***explicar***. La diferencia entre predecir y explicar es análoga a la que hay entre "leyes" y "teorías". Una "ley" es un postulado inductivo de cómo se comportan hechos y fenómenos. Las leyes suelen representarse bastante bien por medio de fórmulas matemáticas, ya que describen patrones de comportamiento observados y predicen comportamientos futuros. Sin embargo, las leyes no explican el porqué de los comportamientos. Por ejemplo, las leyes de los gases o las leyes de masas en química describen y predicen de modo consistente el comportamiento de los gases en una mezcla o la forma como se combi-

Tabla 2.2 Tipos de problemas de atención que generan preguntas

1.	**Frecuencia: incidencia y prevalencia de condiciones.** Mortalidad.
2.	**Factores de riesgo: frecuencia y asociación. ¿Están** involucrados en la cadena causal?
3.	**Diagnóstico: probabilidades; preexamen;** desempeño de pruebas diagnósticas.
4.	**Historia natural. Curso clínico y pronóstico:** distribución de factores pronósticos, asociación y causalidad.
5.	**Eficacia, efectividad y eficiencia de intervenciones** preventivas, terapéuticas o de rehabilitación.
6.	**Consecuencias económicas de la enfermedad,** y su manejo.
7.	Funcionalidad y calidad de vida.

nan cantidades fijas de elementos químicos para formar compuestos, pero no explican por qué suceden los fenómenos descritos. Por el contrario, las teorías (que también son postulados inductivos —probabilísticos—) pretenden explicar por qué y cómo ocurren los fenómenos de interés. La teoría atómica ofrece explicaciones sobre cómo interactúan los elementos, las cuales son consistentes con lo descrito (y lo predicho) por la ley de masas en combinaciones químicas. De hecho, la teoría es más incluyente, por cuanto en su explicación incorpora la ley. No obstante, una ley puede describir, predecir y, en general, funcionar bien sin que sea necesario conocer la explicación subyacente.

En la investigación clínica y epidemiológica pueden reconocerse, entonces, tres tipos de intencionalidad: a) describir o caracterizar un fenómeno o una situación, b) predecir eventos, dado que se determinan las condiciones iniciales, independientemente de la razón por la cual ocurren, y c) explicar (y continuar siendo capaz de predecir) por qué ocurren los eventos.

Las preguntas de investigación pueden, pues, clasificarse, de una manera inicial, de acuerdo con la intencionalidad de la investigación, así:

- Descriptivas: **Buscan reunir información para** describir, enumerar caracterizar, cuantificar y calificar. Estas son preguntas referidas a qué ocurre, cuándo ocurre, a quién, dónde, en qué cantidad o en qué magnitud. Los estudios de frecuencia son típicos ejemplos de las respuestas a tales tipos de preguntas.
- Inferenciales: **A su vez, con dos énfasis diferentes:**
  a. **Pragmático: ¿Cuáles son las condiciones** iniciales asociadas a un desenlace en particular? Por ejemplo, ¿la frecuencia de accidentes de tránsito letales cambia con el uso del cinturón de seguridad?; si es así, ¿la probabilidad de sobrevivir a dichos accidentes es mayor en quienes usan cinturón de seguridad que en quienes no? Nótese que la respuesta es útil para tomar decisiones (recomendar el uso del cinturón de seguridad), independientemente de que se entienda por qué y cómo protege el cinturón.
  b. **Explicativo: Es necesario no solo** establecer si hay asociación entre las condiciones iniciales y los desenlaces, sino también en cómo y por qué ocurren. Tomando el ejemplo anterior, un estudio pragmático podría encontrar que la frecuencia de accidentes letales, de hecho, es mayor en personas que usan el cinturón de seguridad que en quienes no lo usan, y la utilización acrítica de tal información llevaría a recomendar que no se use el cinturón. Un estudio explicativo podría sugerir que cuantas veces se involucre la misma cantidad de energía cinética, el trauma en quienes usan el cinturón de seguridad es menor, pero también, hallar que los vehículos en los cuales los pasajeros usan el cinturón transitan a mayor velocidad, por una sensación de seguridad que experimenta el conductor, de suerte que la frecuencia de accidentes y la cantidad de energía cinética involucrada son mayores. Esta ***explicación*** **permite describir el** comportamiento observado al entender la causa aparente del fenómeno, e identificar maneras de modificarlo. La idea no sería suprimir el cinturón, sino modificar la conducta de quienes lo usan, para que transiten a velocidades razonables.

Independientemente de si son descriptivas o inferenciales, las preguntas también pueden clasificarse según como se enfoquen: en cuanto a variabilidad (***cuantitativas***) **o en cuanto a** procesos (***cualitativas***).

Las preguntas enfocadas en variabilidad involucran la cuantificación de variables, con el fin de determinar frecuencias y contrastarlas, o de establecer asociaciones o correlaciones. Pertenecen primariamente al entorno de la investigación cuantitativa. En general, son el punto de origen de un proceso de investigación, y condicionan su estrategia y su metodología. Por ejemplo, las preguntas relacionadas con la asociación entre un tratamiento y la probabilidad de curación (eficacia-efectividad) apuntan al uso de estrategias experimentales o metaanálisis de experimentos como la forma más apropiada de contestarlas.

Las preguntas enfocadas en procesos suelen pertenecer más al ámbito de la investigación *cualitativa.* **Intentan caracterizar por qué y** cómo, más que determinar asociación y causalidad. Ayudan a ***comprender*** **los fenómenos** observados, más que a ***explicarlos*** **o** ***predecirlos***

en términos de causa-efecto o asociación de variables predictoras y desenlaces.

Utilizando el enfoque epistemológico de los tres mundos de Popper, se puede afirmar que el ámbito de las preguntas enfocadas en variabilidad conciernen más a menudo a los entes y a los fenómenos de los mundos 1 y 3, mientras que el dominio natural de las preguntas con contenido cualitativo está en el mundo 2 (la experiencia subjetiva).

Formulación de preguntas de investigación adecuadas

No todas las preguntas de investigación tienen igual importancia o similares probabilidades de conducir a una respuesta exitosa. Formular preguntas de investigación adecuadas es una parte clave del éxito de una investigación. Sin embargo, formular buenas preguntas no es un proceso estandarizado uniforme y fácil de enseñar o de transferir. El ejercicio de formulación es una destreza que tiene que ver con el talento natural y la inspiración, y, sobre todo, con mucha dedicación y práctica. Sin embargo, es posible, como en el cultivo de cualquier habilidad, desarrollar y utilizar técnicas que permitan mejorar el desempeño. Ante todo, es necesario identificar los elementos que caracterizan una buena pregunta.

Atributos deseables en las preguntas de investigación

Una buena pregunta de investigación debe ser:

Importante: La pregunta de investigación debe ser importante en varios aspectos: la respuesta añade, complementa o mejora el conocimiento acerca del problema específico, y el problema específico es relevante para la disciplina o el área profesional involucrada y para los usuarios del conocimiento producido (pacientes, clínicos, otros investigadores). Se puede evaluar este aspecto de la pregunta con el test del "¿Y qué importa si…?" Si la pregunta no se hace, ¿qué consecuencias tiene para la comprensión del problema, para su solución, para los pacientes, etc.? Por supuesto, la importancia se debe juzgar en relación con un entorno definido y con otras preguntas que pueden competir por el tiempo, el interés del investigador y los recursos disponibles. Otros aspectos relacionados con la importancia son sus implicaciones políticas, culturales y económicas, y si hay involucrados aspectos éticos y de equidad.

Interesante: El interés tiene que ver con qué tan motivadora es la pregunta para el investigador, porque le resulta intrigante y retadora (curiosidad); también, con el hecho de que la respuesta sea importante (ver párrafo anterior), y con sus aspectos innovadores.

Suficientemente específica: La especificidad es necesaria para que sea factible encontrar una respuesta que signifique algo concreto.

Suficientemente amplia: Debe ser útil para teorizar y articularse con el cuerpo de conocimientos en torno al problema de interés.

Contestable: Que se la pueda abordar con los recursos y los medios disponibles para el investigador. En este punto, es bueno llamar la atención sobre los llamados errores tipo III y tipo IV. El error tipo III se enuncia como el hecho de hacer la pregunta equivocada. A este respecto, es claramente preferible tener una respuesta aproximada a la pregunta correcta, y no tener una respuesta precisa a la pregunta equivocada. El tipo IV es hacerse una pregunta que no vale la pena contestarse.

Éticamente apropiada: El último atributo en la lista, pero de ninguna manera el último en jerarquía.

Estrategias para la formulación de preguntas

Los pasos para formular preguntas son: 1) identificar y analizar el problema, 2) buscar información adicional, 3) enfocarse en aspectos específicos, 4) formular las preguntas, 5) evaluarlas, 6) escoger la estrategia de investigación y 7) formular los objetivos; y, finalmente, 8) revisar iterativamente el ciclo.

- Identificación y caracterización del problema general:
  El primer paso consiste en identificar el problema de interés, que en el ámbito de los cuidados en salud suele estar relacionado con la toma adecuada de decisiones (diagnósticas, terapéuticas, etc.). La visión del problema suele ser inicialmente vaga y pobremente definida, y puede parecer confusa por la presencia de múltiples facetas o aspectos. Un buen ejercicio es intentar una reducción del problema a sus componentes y describir sus relaciones; es decir, ***analizar*** el problema. Al identificar sus componente y aventurar un esquema de sus relaciones, se

identifican los puntos en los cuales se necesitan respuestas para alimentar los procesos de decisión.

- Buscar información adicional:
  En este punto es conveniente revisar someramente la literatura científica, comentar con colegas y expertos y revisar, a la luz de esa información, el análisis preliminar. Generalmente, a partir de ese punto hay aspectos que se empiezan a perfilar como de particular relevancia en la solución del problema.
- Enfocarse en aspectos específicos y convertirlos en preguntas:
  En este punto se escogen algunos de los aspectos del problema y se empieza a decidir sobre los tipos de preguntas que es adecuado formular al respecto: ¿Se necesita información descriptiva, o la respuesta que se quiere buscar es una inferencia? ¿Dicha inferencia será de tipo pragmático o explicativo? ¿El aspecto de interés es determinar o contrastar variabilidad (cuantitativo), o es comprender procesos (cualitativo)?
- Formulación de la pregunta:
  Según el tipo escogido de pregunta, las estrategias de formulación pueden variar. Independientemente del tipo de pregunta, el proceso de formulación implica la escogencia y el uso de términos que definan conceptos, y luego, la operacionalización de los componentes de la pregunta. Operacionalizar quiere decir identificar las *operaciones* necesarias para medir los componentes (responder la pregunta). Vale la pena recordar que la medición puede ser cuantitativa (evaluar magnitudes) o cualitativa (clasificar en categorías, caracterizar, describir, narrar).

Preguntas descriptivas: Para preguntas de carácter cuantitativo descriptivo, se utilizan pronombres tales como ***qué, cuál, quién,*** *cuánto, cuándo*…, que se refieren a la característica que se quiere describir y cuantificar. Después se debe definir la población (en quiénes), qué es lo que se va a describir y su marco en relación con el tiempo (cuándo), el lugar (dónde) y la persona (características específicas que definen la pertenencia a la población; por ejemplo, edad, sexo, raza, nivel socioeconómico, presencia de condiciones clínicas, etc.).

EJEMPLO

El problema general es el aparente aumento de la frecuencia de embarazo en adolescentes. El análisis del problema muestra varios componentes relacionados con el aparente aumento, y el investigador genera conjeturas preliminares: puede haber un verdadero aumento en la población general, o puede ser que esté cambiando el tipo de población que está consultando a la institución de interés, lo cual resulta en un aumento aparente. Antes de preocuparse por las causas del aumento, pensar en comprender y explicar el porqué, determinar los factores asociados y pensar en sistemas de prevención, es necesario determinar si en realidad hay un aumento de la tasa de embarazos en adolescentes o si se trata de un artificio. En este punto se recoge información adicional que, eventualmente, puede aclarar la situación y hacer que el investigador se enfoque en otro aspecto del problema. Pero puede ser que no ocurra esto, y entonces debe plantear preguntas de tipo descriptivo cuantitativo: ¿Cuál es la frecuencia de embarazos en adolescentes? ¿Se ha incrementado recientemente?

Estas preguntas aún no están lo suficientemente enfocadas ni son tan específicas como para que su respuesta contribuya a la solución del problema. Se debe, entonces, hacer el ejercicio de refinar el planteamiento de la pregunta con los elementos propuestos:

a. **Concepto: Embarazo en adolescentes.**
b. **Población:**
   - Personas: Adolescentes, rango de edades, estado civil, actividad, sistema de seguridad social, etc.
   - Tiempo: fechas (calendario) e intervalo de interés (por ejemplo, un año).
   - Lugar: Ciudad, región, etc.
c. **Operacionalización: definición *operativa* de** embarazo (por ejemplo: motivo de consulta, test de embarazo, etc.), adolescente.

Las preguntas de interés podrían ser:

a. **Del total de adolescentes de sexo femenino** (definidas como…) en la población escolar de los colegios XXX, ¿en cuántas se determinará la presencia de embarazo (medido como…) en un período de (XX tiempo) a partir de (XX fecha)?
b. **¿Cuál fue la frecuencia de embarazo en la** misma población y con las mismas definiciones y los mismos sistemas de medición en un período de (XX tiempo) durante los años…?

Preguntas inferenciales: En el caso de preguntas inferenciales, es aconsejable tener en cuenta cinco aspectos: a) dominio de la pregunta (riesgo, pronóstico, curso clínico, tratamiento, examen diagnóstico, consecuencias económicas, causalidad, calidad de vida, etc.); b) población de interés: quién, definido en tiempo, lugar y persona; c) factor de exposición: expuesto a (medicamento, factor de riesgo, examen diagnóstico…); d) exposición de control, y e) desenlace. Como en el caso anterior, es necesario operacionalizar los conceptos y las variables que se van a medir.

EJEMPLO

El problema general surge a la hora de decidir si vale la pena tratar con amoxicilina a un niño con un episodio de otitis media aguda. En muchos países, los médicos están manejando a los niños con diagnóstico de otitis media aguda no complicada sin administrar antibióticos, y los padres de algunos pacientes han manifestado su deseo de que sus hijos no reciban antibióticos. El análisis del problema evidencia varios puntos de interés en la toma de decisiones. El punto más importante se relaciona con el curso clínico y el pronóstico de la otitis media si no se recibe antibiótico, y si administrar amoxicilina es ventajoso para el paciente, en comparación con no darle antibióticos. Al recogerse información adicional, se aprecia que hay estudios realizados en países europeos, y los cuales muestran que la resolución de la otitis media, el tiempo hasta la mejoría del dolor y la frecuencia de complicaciones supurativas son similares en grupos con y sin tratamiento antibiótico. Por otra parte, aunque se trata de un medicamento seguro, su uso no está exento de riesgos, y, por supuesto, implica costos directos. La información adicional también muestra que en muchas ocasiones la etiología de la otitis media aguda no es bacteriana. Si en muchas poblaciones la causa es predominantemente no bacteriana, eso podría explicar la falta de diferencia en cuanto a resultados entre el tratamiento con el antibiótico y sin él. El problema de tomar decisiones tiene ahora un aspecto adicional: la población de interés puede tener una frecuencia de etiología bacteriana mayor que las poblaciones de donde proviene la evidencia, y en ese caso el uso de amoxicilina podría ser ventajoso. En este punto se puede proceder a plantear la pregunta de investigación.

En los niños colombianos con otitis media aguda, ¿es eficaz y seguro administrar amoxicilina? La respuesta a esta pregunta es de tipo pragmático, ya que no explicaría por qué la eficacia y la seguridad podrían ser diferentes de las halladas en otros países, pero permitiría decidir si para la población de interés es apropiado el tratamiento con amoxicilina.

Una pregunta de carácter explicativo podría ser más compleja: ¿Cuáles son la eficacia y la seguridad del tratamiento de la otitis media aguda con y sin evidencia de etiología bacteriana respecto a no dar antibióticos? La respuesta a dicha pregunta permitiría no solo determinar la utilidad de la amoxicilina, sino explicar la causa de su éxito aparente o de su fracaso aparente.

Tal como se explicó a propósito de la formulación de preguntas descriptivas, es necesario refinar la pregunta y operacionalizarla antes de que sea útil como generadora de un proceso de investigación. En el caso de la pregunta pragmática, los elementos serían:

a. **Dominio: efectividad y seguridad de una** intervención terapéutica.
b. **Población de estudio: niños definidos en cuan**to a edad, estado de salud previo y con diagnóstico de otitis media aguda, definida según los criterios para el diagnóstico (por ejemplo, otoscopia, neumotimpanoscopia, timpanocentesis, etc.), de etiología (no determinada, determinada por cultivo, etc.), que reciben cointervenciones (por ejemplo, analgésicos, descongestionantes, irrigación nasal, etc.).
c. **Expuestos a: amoxicilina (presentación,** dosis, vía, duración del tratamiento).
d. **Comparados con: el mismo tipo de pobla**ción, pero expuesta a placebo.
e. **Desenlace: específico: curación (definida** como…), duración de los signos o los síntomas, frecuencia de complicaciones, secuelas, etc.

La pregunta pragmática podría ser: ***Población de estudio:*** **En niños previamente sanos,** con edades entre los 3 meses y los 5 años y con primer episodio de otitis media aguda (definida como diagnóstico clínico por timpanoscopia) de menos de tres días de duración, y que reciben tratamiento con acetaminofén, irrigación nasal y vasoconstrictor tópico (***exposición***), **¿la** administración de amoxicilina oral a dosis de 50 mg/kg/día por 7 días (***desenlace***) **se asocia** a una tasa de curación mayor (definida como tiempo hasta la desaparición de la fiebre) (***exposición de control***) **que cuando se administra** placebo?

- Evaluación de las preguntas:

Después del proceso de focalización y refinamiento, es apropiado volver a revisar si las preguntas tienen los atributos deseados. El acrónimo FINER sirve para recordar los ítems que deben ser evaluados: Factible, Interesante, Novedosa, Éticamente apropiada y Relevante (importante).

- Planteamiento de objetivos y escogencia de la estrategia
  Ayudado por la pregunta, el investigador puede definir los objetivos y escoger la estrategia (diseño) de investigación más adecuada. En este punto, puede descubrir que hay inconvenientes, relacionados con los objetivos o la estrategia, que pueden comprometer la investigación, o bien, el ejercicio de formulación de objetivos le indica la necesidad de modificar las preguntas de investigación. Durante todo el proceso de diseño de la investigación se debe seguir este procedimiento iterativo, donde los pasos previos se revisan a la luz de los desarrollos subsecuentes.

PLANTEAMIENTO Y USO DE HIPÓTESIS EN INVESTIGACIÓN CIENTÍFICA

Las preguntas de investigación relacionadas con la intención de hacer inferencias (predictivas o explicativas) involucran el planteamiento tácito o explícito de hipótesis. Una hipótesis es una suposición sobre las características y las propiedades de un evento, un proceso o una interacción. Es la formulación explícita y organizada de una conjetura acerca de cómo es la naturaleza del evento, del proceso o de la interacción, y aunque son muchas las afirmaciones que se pueden deducir de una observación, la tarea del investigador consiste en escoger la más plausible entre diversas alternativas, y esto puede depender, en gran medida, de la forma como se planteen las hipótesis que se van a evaluar. Una ***pregunta científica inferencial*** puede verse como una *hipótesis* planteada en términos interrogativos, mientras que una hipótesis puede entenderse como los contenidos de una pregunta inferencial planteada en términos asertivos. Afirmar esto constituye una aproximación inicial a la definición de hipótesis científica, pero no es del todo preciso, según se aclarará en la discusión que sigue.

Una de las formas más utilizadas de estructurar una explicación científica es generar un argumento deductivo (véase el apartado "Argumentación lógica") cuya conclusión sea el evento que se quiere explicar. Algunas de las premisas de dicho argumento son afirmaciones acerca de las circunstancias precedentes: los llamados ***datos antecedentes*** (que indican las circunstancias en las cuales se da el hecho que se quiere explicar), mientras la restante es la ***hipótesis científica,*** que vincula dichas circunstancias con el desenlace especificado en la conclusión. Las predicciones científicas tienen exactamente la misma estructura.

En este modelo deductivo-nomológico (*nomos* = ley), la conclusión deducida (el evento predicho) solo será cierta si todas las premisas son ciertas. Los datos antecedentes son especificados por el investigador, de manera tal que *deben* ser ciertos. Según como se verá más adelante en un ejemplo, estos datos antecedentes describen las características de los sujetos y las variables asociadas a los desenlaces que ya han sido estudiadas, y cuyos efectos se quiere aislar o controlar. La calidad de la explicación depende, entonces, de que el argumento lógico sea valedero, y de la medida en la que la hipótesis refleje la verdadera relación de los hechos. (Modelo de C.G. Hempel y P. Oppenheim, basado en el empirismo lógico).

EJEMPLO

Continuando con el problema de investigación sobre el tratamiento con antibióticos de la otitis media, el argumento lógico deductivo se plantearía así:

a. Premisas:
- Datos antecedentes: 1) Se trata de niños previamente sanos (definidos en tiempo, lugar y criterios de inclusión y de exclusión) con otitis media aguda (definida operativamente de cierta manera y medida sin error), y que reciben manejo con analgésicos, irrigación nasal y vasoconstrictora nasal, 2) La amoxicilina a dosis adecuadas mata *in vivo* las bacterias susceptibles.
- Hipótesis: La infección por bacterias susceptibles a la amoxicilina es responsable de la mayoría de las otitis medias.

b. Conclusión: La mayoría de niños que reciben amoxicilina mejoran más rápidamente.

Este modelo presenta varios problemas. En primer lugar, presupone que los datos antecedentes ***tienen que*** ser ciertos, circunstancia difícil de lograr y, muy seguramente, imposible de demostrar: En el ejemplo no puede garantizarse que algunos de los datos antecedentes sean ciertos: algunos de los niños diagnosticados pueden no tener otitis media, no todos reciben de forma homogénea los tratamientos adicionales, y la susceptibilidad de poblaciones de bacterias tiene variaciones, de suerte que no todas las bacterias susceptibles se exponen a la misma dosis de amoxicilina ni reaccionan de la misma manera.

En segundo lugar, la determinación empírica de que la conclusión se cumple (el evento predicho ocurre) no puede ser utilizado como evidencia de la veracidad de la hipótesis, pues el hecho de que se produzca el evento predicho (caso particular) no garantiza que cuantas veces se repita la observación va a ocurrir lo mismo. Karl Popper explicó cómo este tipo de verificación se fundamenta en inferencias inductivas, y lo hizo con el ejemplo de los cisnes blancos: se plantea la hipótesis que ***todos*** los cisnes son blancos. No importa cuántos cisnes blancos se encuentren, esto no comprueba la hipótesis, pues siempre existe la posibilidad de que alguien encuentre un cisne negro. Afirmar que, tan solo por haberse observado una gran cantidad de cisnes blancos y nunca otro de color diferente, los cisnes solo pueden ser blancos es inferir de lo particular una ley general (inducción), lo que constituye una falacia formal conocida como la falacia de *afirmación de la consecuencia.* Por el contrario, hallar un solo ejemplo de cisnes negros permite con certeza (dentro de los límites del error de medición; es decir, de la probabilidad muy pequeña, pero diferente de cero, de confundir un cisne blanco con uno negro) concluir que al menos una de las premisas es falsa.

En resumen, una hipótesis no puede ser *comprobada* o *demostrada* con certeza, pero sí puede ser *refutada* o *falsada*. Esta propuesta de Popper (refutacionismo puro o ingenuo) es la base del método hipotético deductivo de generación de explicaciones científicas, el cual se describirá más adelante. Hay una versión sofisticada del refutacionismo que reconoce que, debido al principio de incertidumbre (el error aleatorio inherente a todo acto de medición), es imposible refutar de forma absoluta una hipótesis basándose en mediciones empíricas.

A pesar de estas dificultades epistemológicas, se puede ver con claridad la importancia de plantear hipótesis en la generación de conocimiento científico. Es necesario ahora mirar con más detalle los procesos de generación y evaluación de hipótesis.

El origen de esa conjetura que se usa como premisa (es decir, de la hipótesis) es similar al de las preguntas científicas mencionadas anteriormente. Sin embargo, a diferencia de las preguntas, la hipótesis es una posible respuesta inventada por el investigador a partir de sus observaciones, de su análisis del problema y de su propio pensamiento creativo.

No toda conjetura funciona igualmente bien como hipótesis científica. Y no toda hipótesis bien planteada proporciona una adecuada explicación científica de los hechos. En este punto es necesario definir qué es una buena hipótesis, por un lado, y, por otro, reflexionar sobre la manera como se juzga si una hipótesis funciona como una buena explicación científica.

Características de una buena hipótesis

En primer lugar, las buenas hipótesis deben tener los mismos atributos considerados deseables en una buena pregunta de investigación: la importancia del tópico o el problema, además del hecho de ser interesantes, innovadoras y éticamente adecuadas, y la factibilidad de su estudio.

Además de lo anterior, hay unos atributos específicos de las buenas hipótesis científicas. La propiedad más importante de una buena hipótesis es su ***refutabilidad;*** es decir, que exista algún medio práctico para poder demostrar que es incorrecta. A veces, pensamos que el papel de un investigador es confirmar una hipótesis, más que refutarla. Sin embargo, son más fuertes (científicamente) las hipótesis fácilmente refutables, pero que sobreviven a repetidos intentos rigurosos de contradecirlas. Para ser refutables, deben reunir dos características: 1) que tengan *contenido empírico* y 2) que sean *restrictivas.*

Por contenido empírico se entiende que sus predicciones (las conclusiones que se deduzcan del argumento lógico del cual son parte) se traduzcan en fenómenos observables y mensurables, y, que, por tanto, caigan en el dominio de las ciencias empíricas.

Las buenas hipótesis son ***prohibiciones;*** mientras más restrictivas sean, mejores son.

En otras palabras, mientras más precisamente sea definida una conjetura, su refutación es más factible y más útiles son los resultados de dicho intento de refutación. Por ejemplo, podríamos plantear la hipótesis de que cierta reacción adversa ocurre cuando dos drogas, A y B, son administradas al tiempo. Si observamos una reacción similar cuando un paciente no recibe A ni B, o, tal vez, B y otra droga C, ¿qué podemos concluir? No podemos refutar nuestra hipótesis. Si, en cambio, hubiéramos formulado nuestra hipótesis como "solo se produce esta reacción adversa cuando se administran conjuntamente las drogas A y B", entonces, la nueva evidencia nos habría forzado a abandonar la primera hipótesis y a formular una nueva, más amplia.

Evaluación de hipótesis

De acuerdo con la discusión previa, es imposible probar la veracidad de una hipótesis científica. Sin embargo, sí es posible estimar su verosimilitud, y, especialmente, comparar diferentes hipótesis involucradas en explicaciones alternativas para un mismo fenómeno. Los criterios usualmente empleados para evaluar la verosimilitud, o ***bondad de ajuste,*** de una hipótesis a los hechos o los fenómenos incluyen los siguientes:

a. **Relevancia:** **Al menos debe ser posible utilizar** la hipótesis como una de las premisas de un argumento deductivo válido cuya conclusión sea el evento que se busca explicar o predecir. También, en principio, se espera que la mejor explicación involucre la hipótesis cuya relación con el evento sea más aparente. Por supuesto, en ocasiones hay hipótesis que en un momento dado parecen ser menos relevantes, pero cuando se amplía el conocimiento se descubren hechos que aumentan su relevancia. Por ejemplo, antes de postularse la teoría microbiana del origen de las enfermedades infecciosas, la relevancia del lavado de las manos para prevenir la fiebre puerperal era poco aparente, pero una vez se interpretó al lavado de las manos como capaz de remover las bacterias causantes de la enfermedad, se comprendió su relevancia como explicativa de la prevención de la contaminación cruzada entre madres puérperas causada por los médicos que las examinaban.

b. **Evaluabilidad:** **Debe existir una manera de** recolectar evidencia que permita evaluar la hipótesis; en particular, que permita intentar su refutación (***falsabilidad;*** ver el apartado "Características de una buena hipótesis").

c. **Compatibilidad:** **La hipótesis debe encajar** correctamente en el conocimiento científico valedero previo. Cuando tal condición no se da, es lícito suponer que, en principio, la hipótesis no es una buena explicación científica. Sin embargo, este criterio debe manejarse con mucho cuidado, pues las verdaderas revoluciones científicas que implican la modificación sustancial (e, incluso, el abandono) de hipótesis aceptadas previamente, o llegar hasta al cambio de paradigmas, ocurren por la aparición de nuevas hipótesis sin consistencia con las precedentes, pero que explican mejor la realidad al concordar mejor con los datos empíricos.

d. **Poder predictivo:** **Una hipótesis robusta suele** ser capaz de explicar de forma satisfactoria un espectro de eventos más amplio que el que le dio origen. La Ley de la Gravitación Universal de Newton es un ejemplo de una hipótesis con gran poder predictivo, por cuanto explica de manera adecuada tanto la mecánica celeste como la caída de objetos en la Tierra. En el extremo opuesto, el del pobre poder predictivo, están las hipótesis ***ad hoc*****, generadas para explicar la ocurrencia** específica de un evento particular.

e. **Simplicidad:** **Las mejores hipótesis, generalmente,** tienen una estructura poco intrincada, y, en principio, si se puede apelar a una sola causa para explicar de forma satisfactoria un fenómeno, dicha explicación es probabilísticamente más factible que la coincidencia de muchas causas. Este es un fenómeno con el cual los clínicos están bastante familiarizados. Cuando un paciente tiene una constelación de síntomas y signos, las hipótesis diagnósticas más plausibles (y con las que un buen clínico suele iniciar su análisis) son las relacionadas con una sola enfermedad. Es posible que un paciente con disnea y edemas en los miembros inferiores tenga una neumonía y un síndrome nefrótico, pero la mayoría de los clínicos pensaría primero (y tendría más probabilidad de acertar) en una sola causa que los explique ambos fenómenos, como, por ejemplo, una insuficiencia cardíaca congestiva que esté produciendo edema pulmonar y congestión pasiva crónica. El principio de la simplicidad

es conocido entre los científicos como la *Cuchilla de Ockham*. De nuevo, se entiende que este principio debe ser aplicado con reservas, pues en ocasiones la explicación más adecuada es la más compleja. En otras palabras, es claramente preferible una hipótesis compleja, pero adecuada, a una simple, pero insuficiente. Sin embargo, hay una belleza y un atractivo profundos en la simplicidad y la elegancia de muchas de las grandes hipótesis científicas.

¿Preguntas de investigación o formulación de hipótesis?

Suele ser más fácil generar preguntas que hipótesis, y, en realidad, se puede entender el planteamiento de una hipótesis como la formulación de una pregunta inferencial en términos asertivos, más que de interrogación.

Para ilustrar este punto, es conveniente examinar un ejemplo: el curso clínico de la nefropatía diabética. Los pacientes con diabetes (sobre todo, si tienen hipertensión arterial concomitante) tienen un elevado riesgo de desarrollar una nefropatía diabética que puede progresar en etapas sucesivas hasta desembocar en insuficiencia renal crónica terminal (IRCT). Se ha establecido que si hay un estricto control de la glucemia, un manejo adecuado de dislipidemia (si ella existe), un estilo de vida saludable (sin sedentarismo, sin tabaquismo, controlando el sobrepeso) y un manejo agresivo de la hipertensión arterial concomitante, el riesgo de progresión de la nefropatía diabética disminuye. También se ha observado que, en el caso de los diabéticos tipo 1, el uso de antihipertensivos que inhiban la enzima convertidora de angiotensina I (IECA) ejerce un efecto nefroprotector adicional al obtenido con el simple control adecuado de la tensión arterial. Hay otra clase de medicamentos antihipertensivos, los bloqueadores de los receptores de la angiotensina II (BRA II), que podrían ejercer un efecto nefroprotector similar, y, de hecho, hay evidencias a favor de dicho efecto en pacientes diabéticos tipo 2 hipertensos. Ambas clases de medicamentos pueden ser, por tanto, eficaces. Sin embargo, el grado de nefroprotección puede variar en cada persona, y los riesgos y los costos asociados a cada clase de medicamento son distintos. ¿Cómo puede el clínico escoger adecuadamente cuál clase de medicamento le ofrecerá a su paciente? La respuesta debe provenir de una investigación capaz de contestar dicho interrogante.

La hipótesis para tal estudio podría plantearse así: "En pacientes diabéticos tipo 2, hipertensos, sin daño de órgano blanco, y que tienen un adecuado control de su glucemia y de otras condiciones comórbidas (por ejemplo, dislipidemia), el tiempo de supervivencia libre de IRCT es mayor, y los riesgos y los costos médicos directos asociados a la terapia antihipertensiva son menores si se les administra un IECA, en vez de un ARA II". Esta conjetura podría haberse planteado también en forma de pregunta de investigación, y contendría los mismos elementos: "En pacientes diabéticos tipo 2 hipertensos sin daño de órgano blanco, y que tienen un adecuado control de su glucemia y de otras condiciones comórbidas (por ejemplo, dislipidemia), ¿qué clase de medicamentos antihipertensivos (IECA o ARA II) es la mejor alternativa de tratamiento, en términos de tiempo de sobrevida libre de IRCT y de riesgos y de costos médicos directos asociados a la terapia antihipertensiva?

Pensar en términos de hipótesis o de pregunta adecuadamente formuladas parece ser equivalente. Sin embargo, el uso de hipótesis ofrece algunas ventajas en la aproximación al problema del planteamiento de un tema de investigación.

Cuando se escoge pensar en términos de hipótesis, se hace necesario plantear de manera explícita las hipótesis alternativas y las consecuencias específicas que se pueden deducir de la conjetura planteada. Las hipótesis alternativas son varias, y corresponden a las combinaciones de desenlaces mutuamente excluyentes que es factible observar, diferentes del predicho por la hipótesis inicial. En el ejemplo que se viene desarrollando, habría tres familias de hipótesis alternativas: 1) la primera, en la que el tiempo de sobrevida libre de diálisis es mayor si se administra un IECA, pero los riesgos y los costos cambian; 2) la segunda, en la que la efectividad del ARA II sea mejor, y 3) la tercera, en la que las efectividades de ambas clases de medicamentos sean equivalentes.

En cada familia de hipótesis alternativas pueden plantearse varias combinaciones de desenlaces mutuamente excluyentes. Así, siguiendo con el ejemplo planteado, en caso de que la efectividad del IECA sea mayor aparecen las siguientes hipótesis alternativas terminales: 1) el tiempo libre de IRCT es mayor si se administra

un IECA, pero los riesgos y los costos médicos directos son mayores; 2) el tiempo libre de IRCT es mayor y los riesgos son menores si se administra un IECA, pero los costos médicos directos son mayores; 3) el tiempo libre de IRCT es mayor y los costos médicos directos son menores si se administra un IECA, pero los riesgos son mayores, y 4) el tiempo libre de IRCT es mayor y los riesgos y los costos médicos directos son equivalentes.

Pensar en términos de hipótesis y hacer el ejercicio de plantear todas las hipótesis alternativas permite deducir e identificar las consecuencias observables en la realidad, en caso de que la hipótesis sea correcta, y qué tipo de resultados permiten refutar la hipótesis formulada. Al anticipar los posibles resultados del estudio, se tiene una idea clara, desde el comienzo, acerca de cuáles variables hay que medir y controlar, cuáles desenlaces hay que contrastar y cómo debe ser la interpretación de los resultados obtenidos.

A MODO DE CONCLUSIÓN: EL MÉTODO HIPOTÉTICO-DEDUCTIVO

La generación de hipótesis es una parte importante del proceso científico. Puede, y debe, ser utilizada en todos los pasos de una investigación como una ayuda para pensar clara y lógicamente acerca del objeto de la investigación. En vez de recolectar datos con el ánimo de ver "qué encontramos", los recolectamos con la clara intención de refutar una conjetura. De este modo, la hipótesis dirige la recolección de la información.

Las suposiciones iniciales acerca de un problema suelen originarse como ejercicios inductivos basados en la observación (abducción[1]), o en conocimientos previos (por ejemplo, bioquímicos). Sin embargo, una vez se formula una hipótesis, esta genera la recolección de información adicional. Las hipótesis tempranas son, necesariamente, menos precisas (menos rigurosas) que las que les siguen, pero esas hipótesis tempranas permiten que un gran número de ideas alternativas sean revisadas y abandonadas de forma eficiente. Muchas conjeturas pueden ser contestadas basándose en el conocimiento existente: la literatura o la información recolectada con otros fines. El proceso de investigación debe reservarse para verificar las hipótesis que hayan sobrevivido a todo tipo de ingeniosos intentos deductivos de refutarlas.

El método hipotético-deductivo puede esquematizarse como sigue:

a. Definir el problema: Es importante empezar con una enunciación clara de cuál es el fenómeno que se quiere investigar. Desde este punto se hará patente si se necesita describir (preguntas descriptivas) o se busca inferir. En este último caso, es apropiado seguir con el ejercicio de generar hipótesis.
b. Generar hipótesis preliminares: En esta etapa es apropiado no limitar la creatividad, y no se debe ser muy estricto ignorando conjeturas que puedan parecer irrelevantes. Más adelante se podrán desechar las hipótesis inadecuadas, y es preferible no descartarlas prematuramente, pues se puede prescindir inadvertidamente de una conjetura promisoria.
c. Recolectar información adicional, que puede provenir de datos preexistentes y resultados de otras investigaciones, o puede ser recogida *de novo* con la intención de evaluar aspectos concretos de las hipótesis preliminares. Con estos datos se puede empezar a descartar hipótesis poco promisorias y refinar las supervivientes.
d. Formular una hipótesis definitiva: La hipótesis superviviente, a menudo, es una hipótesis estadística, planteada en términos probabilísticos, que es evaluada utilizando métodos matemáticos.
e. Deducir las consecuencias que se derivan de la hipótesis: Dado que una buena hipótesis es de contenido empírico y de poder predictivo, y que es restrictiva, usualmente es posible deducir las predicciones y las consecuencias esperadas en caso de que la hipótesis sea cierta. Las mejores inferencias deductivas son "prohibiciones" que facilitan recolectar información con el fin de refutarlas. El ejercicio deductivo debe ser riguroso para producir un argumento lógico válido. Si tal condición se produce, los datos empíricos que estén en desacuerdo con una predicción invalidan con certeza la hipótesis (*modus tollens*), dentro de los límites del principio de incertidumbre.

[1] Entre los significados del término "abducción" está "Silogismo cuya premisa mayor es evidente y la menor menos evidente o solo probable" (diccionario de la RAE).

f. **Recolectar información específicamente dirigida** a refutar las predicciones de la hipótesis; es decir, llevar a cabo la investigación.
g. **Analizar los resultados (evaluar las hipótesis).**
h. **Generar conclusiones, preparar y publicar** informes de la investigación.
i. **Aplicar la explicación encontrada (la hipótesis evaluada)**, monitorizar su desempeño y repetir el proceso hasta encontrar una solución satisfactoria al problema. Usualmente, aparecen refinamientos o nuevos aspectos del problema, que obligan a generar nuevas hipótesis y a continuar el ciclo de investigación.

Lecturas sugeridas

Buck C. Popper's philosophy for epidemiologists. Int J Epidemiol. 1975;4:159-68.

Carnap R. An introduction to the philosophy of science. New York: Dover Publications; 1966.

Carter KC. A first course in logic. Londres: Pearson Longman; 2003.

Chalmers AF. ¿Qué es esa cosa llamada ciencia? Buenos Aires: Siglo XXI Editores; 1988.

Dennis RJ, Lozano JM. La pregunta clínica: cómo elaborar preguntas que se puedan contestar. En: Programa de Actualización Médica Permanente. Ascofame; 1998. p. 13-8.

Miller RW. Introduction to logic. Study Guide. Nueva Jersey: Prentice Hall; 1997.

Pérez R. ¿Existe el método científico? México DF: El Colegio Nacional y Fondo de Cultura Económica; 1998.

Popper K. Three worlds. University of Michigan [internet]. 1978 [citado 2014 nov. 10]. Disponible en: https://www.thee-online.com/Documents/Popper-3Worlds.pdf

Popper K. La lógica de la investigación científica. Madrid: Tecnos; 1962.

Richardson WS, Wilson MC, Nishikawa J, et al. The well built clinical questions: a key to evidence-based decisions. ACP J Club. 1995;123:A12-3.

Rothman KJ. Philosophy of scientific inference. En: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008. p. 7-10.

Sackett D. Richardson WS, Haynes RB, et al. How to ask clinical questions you can answer. En: Evidence-Based Medicine: how to practice and teach EBM. Oxford: Oxford University Press; 1997.

CAPÍTULO

3 Búsqueda de la literatura médica

Álvaro J. Ruiz Morales
Óscar Mauricio Muñoz Velandia

INTRODUCCIÓN

El proceso investigativo en medicina empieza con la conversión de una duda clínica o de una observación en una pregunta de investigación. Formular la pregunta implica determinar si existe información suficiente sobre el tema, garantizar que no exista ya una respuesta y asegurar que se tendrá en cuenta toda la información relevante para el proceso de investigación.

La búsqueda de la literatura científica a través de medios electrónicos y en bases de datos facilita el proceso, y ha reemplazado casi por completo los medios impresos. La facilidad y la rapidez de las búsquedas, la opción de obtener de manera inmediata resultados con los artículos en texto completo y la interconexión entre bibliotecas y diferentes bases de datos han aumentado la opción de encontrar la información relevante de forma eficiente.

Sin embargo, la facilidad para la búsqueda y la magnitud de las bases de datos disponibles han creado otro problema, contrario al que se vivía en la época de búsquedas en bases impresas: ahora la dificultad no estriba, principalmente, en encontrar la información relevante (que sigue siendo el objetivo principal de la búsqueda), sino en no encontrar información innecesaria o que no es pertinente. El ideal es encontrar lo que se necesita, todo lo que se necesita y solo lo que es necesario[1].

Por esta razón se requiere tener información amplia y correcta sobre las bases de datos y los sitios de información científica, y es más necesario aún adquirir destrezas para conducir el proceso de búsqueda[2].

Se habla de *sensibilidad de la búsqueda* como la característica de la búsqueda que permite encontrar la información importante que exista. Corresponde a la proporción de artículos relevantes identificados por una estrategia de búsqueda como porcentaje del total de artículos relevantes para el tema.

Y es una excelente medida de la habilidad de quien lleva a cabo una estrategia de búsqueda para encontrar la literatura científica apropiada.

Claramente, cuanto mayor sea el número de revistas científicas contenidas en la base, mejor será la sensibilidad. Como ejemplo, Medical Literature Analysis and Retrieval System Online (MEDLINE), que es la base de datos bibliográfica de ciencias de la salud y

de información biomédica de la Biblioteca Nacional de Medicina de los Institutos de Salud de los Estados Unidos, es una de las dos colecciones existentes más amplias de literatura. Incluye la información bibliográfica de artículos de revistas académicas que cubren áreas como medicina, odontología, enfermería, farmacia, toxicología, medicina veterinaria, salud pública, economía clínica, y, en general, temas sobre atención en salud.

La base de datos, presentada en una versión gratuita en línea (llamada PubMed), contiene en la actualidad más de 21,6 millones de registros, provenientes de 5639 publicaciones desde 1950. Cerca del 50 % de los artículos provienen de los Estados Unidos, pero el 88 % del total son publicaciones en inglés y el 76 % del total tienen resumen en inglés, lo que la hace ampliamente asequible.

La sensibilidad de una búsqueda en PubMed es claramente alta, por el gran número de publicaciones, entre las que es fácil encontrar la información relevante.

Por otra parte, la ***especificidad de la búsqueda*** se refiere a la capacidad para excluir artículos irrelevantes. Es el porcentaje de artículos relevantes identificados mediante una estrategia de búsqueda del total de artículos (relevantes e irrelevantes) encontrados por la estrategia, y es una medida de la habilidad de la búsqueda para excluir artículos irrelevantes.

Es claro que la especificidad de una estrategia tenderá a ser menor cuanto mayor sea la base de datos.

Pero, a diferencia de la sensibilidad, que depende, principalmente, de la magnitud de la información contenida en la base de datos, la especificidad de la búsqueda depende del operador: de su experiencia, del uso de filtros o algoritmos y del tiempo dedicado a la búsqueda.

Para una buena búsqueda, entonces, es necesario disponer de buenas bases, y cuantas más se consulten y mejor sea su calidad, mejor será la sensibilidad. Pero también es necesario tener destrezas que permitan utilizar con eficiencia filtros, algoritmos y trucos para mejorar la especificidad. Esto lo dan únicamente la formación en estrategias de búsqueda y la experiencia.

Así pues, una buena búsqueda necesita dos componentes principales: 1) identificación de buenas bases de datos para hacer la búsqueda, y 2) conocimientos con experiencia para diseñar las estrategias.

La búsqueda apropiada de la literatura le permitirá al investigador:

- Conocer los antecedentes relacionados con el tema de investigación, las estrategias que se han utilizado para abordar el tema, el estado del conocimiento y las brechas de información que requieren mayor estudio.
- Determinar la importancia de la pregunta para la salud de los individuos y para las comunidades, cuantificar la carga de la enfermedad, comparar frecuencias entre comunidades o detectar diferencias regionales, conocer los enfoques diagnósticos o terapéuticos y revisar sus evaluaciones.
- Aprender de la experiencia de otros, de los aciertos y de los errores en el enfoque del tema de interés. También, identificar estrategias de medición y clasificación de desenlaces y otras condiciones que faciliten comparar resultados con los estudios existentes y minimizar los errores metodológicos.
- Analizar los resultados de su investigación a la luz de los conocimientos existentes, identificar las fortalezas y las debilidades del estudio y los aportes nuevos que ofrece.

Una pregunta clínica bien formulada suele aportar los elementos necesarios para una búsqueda eficiente, ***sensible*** (que detecte la mayoría de los títulos relevantes sobre el tema que se quiere revisar) y *específica* (que excluya las publicaciones irrelevantes, de temas diferentes o con diseños que no satisfagan el juicio del investigador).

Para adelantar una búsqueda adecuada, se sugieren los siguientes pasos:

- Seleccionar las palabras clave para la búsqueda a partir de la pregunta clínica.
- Seleccionar la fuente de información apropiada para identificar estudios relevantes.
- Diseñar una estrategia de búsqueda reproducible, sensible y específica.
- Buscar de forma dinámica; repetir el proceso luego de la modificación de los términos, como reacción a los resultados preliminares.

OBJETIVOS DE LA BÚSQUEDA

Se hace referencia a información cuando se obtiene noticia de algo. Sin embargo, la información puede no aportar elementos adecuados o suficientes para que la toma de decisiones clínicas sea apropiada. La información puede estar limitada por los matices propios del informador, y no siempre es balanceada y objetiva.

La evidencia proviene de la información, con un matiz que la diferencia: la calidad de la información, la ausencia de sesgos, el control de errores y de factores de confusión, los métodos correctamente utilizados para la medición y para la recolección, el análisis correcto y otros aspectos permiten concluir si hay, o no, información que se pueda interpretar como certeza clara, manifiesta y tan perceptible que no deje duda sobre su aporte a conocer la verdad.

El objetivo de la búsqueda es identificar, de forma eficiente, la mejor evidencia disponible para contestar una pregunta clínica, y obtener información suficiente para lograr un buen balance en la información[3].

Por ejemplo, si el lector va a adelantar una búsqueda sobre los beneficios de una intervención terapéutica, una buena búsqueda debe brindar información veraz, no solo sobre la efectividad de la intervención, en los aspectos más importantes, sino sobre los riesgos (e, idealmente, los costos) que la intervención conlleva, así como permitir el análisis del balance entre beneficios, por un lado, y riesgos y costos, por otro.

PROCESO DE BÚSQUEDA: PASOS

Una pregunta bien definida, usualmente, incluye precisión en los siguientes aspectos:

- Población o tipo de participantes implicados.
- Tipo de exposición que la persona o la población experimentan (factor de riesgo, intervención o prueba diagnóstica).
- Definición de la *población control*, con la cual se compara la exposición.
- Desenlaces de interés.

A estos cuatro aspectos se los conoce como la *estructura* , que es aplicable en la gran mayoría de preguntas: población, intervención, control y *outcome* (desenlace).

En la búsqueda, también es útil definir el tipo de pregunta según el enfoque que se le quiera dar (si es una pregunta sobre etiología, diagnóstico, tratamiento, pronóstico o guías de predicción clínica), y, si el buscador lo permite, seleccionar el espectro que se quiere (amplio o estrecho).

Cuando se piensa hacer una búsqueda, es útil agregar un *término a la búsqueda*: el diseño de estudio que es más apropiado para contestar la pregunta. Este último término puede hacer la gran diferencia entre hallar o no información o evidencia (definida como información incontrovertible).

En resumen, deben definirse los siguientes aspectos básicos antes de la búsqueda, para facilitarla **(tabla 3.1)**.

Balance entre sensibilidad y especificidad de la búsqueda

La sensibilidad de la búsqueda, como ya se mencionó, se refiere a la capacidad de la búsqueda para identificar todos los artículos relevantes para el tema. Las búsquedas de la literatura imitan el comportamiento de las pruebas diagnósticas, y una búsqueda exhaustiva, que garantice la identificación de todos los artículos relevantes, también identificará muchos artículos irrelevantes.

La especificidad se refiere a la capacidad de la búsqueda para identificar únicamente artículos que se ajusten a la necesidad. Una búsqueda muy específica tiene como ventaja que con ella se evita el desgaste de evaluar artículos irrelevantes, aunque para esto pueda pagarse un precio: perder algunos artículos relevantes.

Tabla 3.1 Aspectos básicos previos a la búsqueda

	Del escenario clínico a la estrategia de búsqueda
Población	Hipertensión arterial ( ); personas mayores de 70 años.
Intervención	Combinación de medicamentos en una tableta.
Comparador	Combinación de medicamentos por separado.
Desenlace	Adherencia.
Tipo de pregunta	Tratamiento.
Tipo de estudio	Experimentos clínicos aleatorizados ( ); metaanálisis.

Fuente: Elaboración propia de los autores.

El investigador debe establecer el balance que quiere tener entre sensibilidad y especificidad, y que le permita obtener la información de forma eficiente. Demasiada sensibilidad producirá un exagerado número de artículos irrelevantes, y una especificidad muy alta hará que se pierdan algunos artículos relevantes.

En términos del balance entre sensibilidad y especificidad de la búsqueda, no hay respuestas correctas, pero sí algunas ideas generales, basadas en el sentido común. Si el investigador adelanta una revisión sistemática de la literatura, deberá procurar que su búsqueda tenga una alta sensibilidad, para no comprometer la validez ni la relevancia de su estudio. Si el investigador, simplemente, busca un artículo para revisar un tema puntual, una búsqueda específica suele ser lo más eficiente.

SELECCIÓN DE LAS PALABRAS CLAVE

Las palabras tomadas de una pregunta pueden ser usadas como palabras clave para una búsqueda provechosa de la literatura biomédica. No obstante, es innecesario usar todos los términos para una búsqueda; usualmente, uno o dos pueden ser suficientes.

El fundamento para las búsquedas en literatura biomédica son los términos que se han estandarizado, como resultado del trabajo hecho por la Biblioteca Nacional de Medicina, que es parte de los Institutos Nacionales de Salud de Estados Unidos (National Library of Medicine, National Institutes of Health, [NLM/NIH]). En busca de estandarización, se hizo un diccionario de términos que describen, de manera coherente y regular, un número muy grande de conceptos biomédicos.

EJEMPLO

Se quiere hacer una búsqueda sobre los beneficios y los riesgos del tratamiento de la hipertensión sistólica aislada en mayores de 65 años. Para la selección inicial de los términos en inglés, el investigador se encuentra ante múltiples opciones para la enfermedad básica que busca: "high blood pressure, hypertension, high pressure, renovascular hypertension, hypertensive disease". La estandarización hecha por la Biblioteca Nacional de Medicina (NLM en inglés), creó un término único, la forma oficial para referirse a la hipertensión arterial: hypertension.

Estos términos estandarizados, conocidos como ***términos MeSH*** (por las siglas en inglés de su nombre: ***medical subject headings***, o encabezados médicos temáticos) corresponden a palabras o frases mediante las cuales se clasifican los temas principales de un artículo. Si los artículos médicos se clasificaran según los términos que cada autor o cada lector considerara apropiados, habría tal diversidad y tan enorme número de clasificaciones heterogéneas que sería difícil identificar los artículos relevantes.

También habría una gran variabilidad si se clasificara el tema de un artículo por palabras tomadas del título o del subtítulo. Por ejemplo, si se quisiera identificar publicaciones que tratan sobre trabajo de parto y se utilizaran estas palabras, la búsqueda se limitaría a la literatura publicada en español, pues las publicaciones en inglés estadounidense vendrían identificados con la palabra ***labor***, y la palabra ***labour*** identificaría a los escritos en inglés ***británico***.

Para buscar una solución y mejorar la sensibilidad de las búsquedas, los Institutos Nacionales de Salud de los Estados Unidos adoptaron una clasificación única y estándar, para cualquier publicación disponible. La clasificación y los artículos clasificados se publican en el Index Medicus o en Medline, independientemente del idioma original del artículo. Así, al buscar el término MeSH "labor", se identifican todos los estudios que clasificaron como relacionados con el trabajo de parto, independientemente del idioma original del artículo.

Resulta conveniente cruzar los resultados iniciales de la búsqueda con el término correspondiente al diseño de estudio. Los términos tomados de una pregunta bien elaborada se incorporan directamente como términos de búsqueda. También pueden incluirse como términos "MeSH". Si no es así, deben buscarse los sinónimos o los términos correspondientes en el MeSH para aumentar la sensibilidad. Todos los servicios de búsqueda que ofrecen Medline y las bases de datos relacionadas tienen sus diccionarios de términos MeSH, los cuales no hay que saber de memoria. Basta con utilizar los diccionarios adjuntos de los programas (MeSH *browsers*) para encontrar la palabra justa y el árbol al cual pertenece la palabra. El diccionario de términos MeSH de PubMed es muy útil, se trata del sitio más ágil, actualizado y amigable de Internet para la búsqueda en Medline (http://www.ncbi.nlm.nih.gov/pubmed), perteneciente

al National Center for Biotechnology Information, de los Institutos Nacionales de Salud de los Estados Unidos. Mediante el uso de los términos MeSH, se puede garantizar una alta eficiencia en la búsqueda, pues su uso está muy generalizado y el lenguaje universal que significa permite evitar los escollos de las diferencias lingüísticas y culturales.

Precisión en la pregunta

EJEMPLO

Siguiendo las indicaciones descritas respecto a la redacción de una pregunta, un investigador quiere averiguar si la vacuna contra la influenza en personas mayores de 80 años que sufran enfermedad pulmonar obstructiva crónica (EPOC) disminuye el riesgo de hospitalizaciones por neumonía. La pregunta fue expuesta de la siguiente manera: "¿Permite la inmunización contra la influenza viral, en adultos mayores de 80 años que sufran EPOC, disminuir el riesgo de hospitalización por neumonía, al menos en el 50 % respecto a una población similar no vacunada?

Para la búsqueda, algunos términos clave se encuentran destacados con letra cursiva. Una forma apropiada de iniciar la búsqueda puede ser utilizar los términos ***influenza viral, neumonía*** y *prevención*. Si se limita la búsqueda a estos términos, se encontrará infinidad de publicaciones que aportan información, pero las realmente útiles, las que ofrecen la evidencia, corren el riesgo de quedar sepultadas en un mar de artículos, y probablemente no lleguen a ser evaluadas adecuadamente por falta de tiempo y fatiga del investigador, quien tendrá por delante una tarea titánica.

Aun si el investigador usa solo un par de términos para adelantar la búsqueda, el hecho de tener la pregunta bien definida le permite identificar y excluir con facilidad los artículos que no son relevantes. Si el investigador incorpora todos los términos en la búsqueda, esta tiende a ser cada vez más específica, pero corre el riesgo de perder un número creciente de artículos relevantes. Usar una estrategia menos específica, pero más sensible, le permitirá seleccionar una lista de títulos de artículos que puede luego depurar, simplemente, revisando los títulos y los resúmenes estructurados según los criterios de la pregunta de investigación.

Tipo de diseño apropiado para la pregunta

Al añadir un diseño metodológico, el más ajustado a las necesidades, a los términos de búsqueda, se seleccionan las publicaciones con mejores posibilidades de responder la pregunta y se eliminarán numerosas publicaciones irrelevantes.

EJEMPLO

Se quiere evaluar la efectividad y la seguridad de la terapia farmacológica en pacientes hipertensos mayores de 70 años. Probablemente, la mejor estrategia para el investigador sea la búsqueda de revisiones sistemáticas de la literatura, o de metaanálisis o de una guía de práctica clínica (GPC) basada en evidencia.

Para la estrategia de búsqueda se utilizan los términos ***hypertension*** y ***elderly***, como términos MeSH mayores, ***drug therapy*** como subencabezado MeSH y ***meta-analysis*** como tipo de publicación. La sintaxis que la página de PubMed genera queda así: (((hypertension[MeSH Major Topic]) AND elderly[MeSH Major Topic]) AND drug therapy[MeSH Subheading]) AND meta-analysis[Publication Type].

Pese a lo anterior, puede suceder que para algunas intervenciones sea difícil encontrar ECA o metaanálisis, bien sea porque no es práctico hacerlos o porque puede ser éticamente cuestionable conducirlos. Por ejemplo, si se quiere evaluar si acostar a los lactantes en posición supina reduce el riesgo de muerte súbita del infante, es posible encontrar evidencia sustancial y de buena calidad en estudios de cohortes de grandes poblaciones, en los que se ha visto una reducción importante en la incidencia de muerte súbita infantil después de iniciar campañas para promover que se acueste a los infantes en decúbito supino. Hacer un experimento clínico al respecto sería muy difícil, dado lo poco común del desenlace y lo difícil que sería controlar variables que podrían confundir la interpretación en un experimento clínico.

Cuando una pregunta evalúa un pronóstico, probablemente el estudio práctico que aporta la mejor evidencia es el estudio de cohortes, pero el estudio ideal para estos casos, no siempre factible por razones logísticas y éticas, es el ECA. Si se quiere identificar factores de riesgo para

condiciones infrecuentes, un estudio de casos y controles puede ofrecer la respuesta. Así, hay muchos otros tipos de estudios que son apropiados para contestar diferentes preguntas.

Generalmente, basta el cruce de dos o tres términos para identificar un número de estudios que sea suficiente para contestar la pregunta de investigación con información de buena calidad. Al incluir demasiados términos, la búsqueda disminuye su sensibilidad, pues no encuentra todos los artículos de relevancia.

Un término puede aparecer referenciado de varias formas diferentes (en especial, cuando no se utilizan los términos MeSH), de manera que incluir más de uno de esos términos puede hacer que la búsqueda sea más sensible. Volviendo al ejemplo, en una búsqueda de artículos que hagan referencia en el título al trabajo de parto, se deben buscar no solo los que tengan la palabra *labor*. Se deben incluir los que tengan el término británico ***labour***. En la medida en que se revisan los artículos hallados, se identificarán los diferentes campos del resumen y las palabras que deben incluirse en los términos de búsquedas secundarias y laterales.

SELECCIÓN DE LA FUENTE DE INFORMACIÓN APROPIADA

Fuentes de información y de evidencia comúnmente utilizadas

Una vez planteada la pregunta clínica de interés, el paso siguiente es buscar la evidencia que permita llegar a la respuesta correcta. Las fuentes usuales de información con las cuales cuenta un clínico pueden resumirse en las siguientes:

La base de datos personal

Las personas suelen recordar (aunque con frecuencia en forma poco clara) dónde leyeron o quién dijo algo acerca de la conveniencia de un tratamiento, la utilidad de un examen diagnóstico o la asociación causal entre una determinada exposición y un desenlace. Generalmente se acude a la biblioteca personal en busca de cualquier ilustración o para refrescar aquello que se cree recordar. Habitualmente, dicha biblioteca es una colección de libros de texto y manuales, con variadas fechas de publicación y diversas metodologías de investigación, y donde comúnmente se incluyen algunos artículos de revisión, artículos originales, documentos obtenidos por Internet, anotaciones de conferencias, tablas y otras fuentes. La calidad y la actualidad de la información contenida ahí pueden variar mucho.

En general, la base de datos personal es buena y actualizada en los temas en los cuales el investigador ha estado más interesado, y también suele reflejar las deficiencias y las debilidades del investigador, puesto que, por ejemplo, habitualmente excluye las publicaciones cuyas metodologías o resultados no lo hacen sentir cómodo.

Si bien esta base suministrará alguna información acerca de la pregunta, es difícil precisar si refleja la existencia de evidencia en la literatura médica, y, a la vez, si esa literatura refleja la disponibilidad de evidencia y presenta de forma balanceada las fortalezas y las debilidades de los estudios que respaldan diferentes puntos de vista.

La opinión de colegas

En medicina, acudir a los colegas con grados variables de experiencia en el tema que se investiga es una estrategia muy corriente; puede ser una forma efectiva de informarse, cuando la asesoría es brindada por un experto en el tema, pero tiene varios inconvenientes: la información suministrada puede no fundamentarse en evidencia científica, no se conoce el proceso de búsqueda de evidencia que se siguió (si se siguió alguno), y puede ser difícil evaluar los fundamentos y la validez de las opiniones expresadas, y las conclusiones pueden discrepar de la evidencia.

Cuando se toma como evidencia la experiencia personal, con frecuencia se les da más relevancia a las situaciones insólitas que a las comunes y se da más importancia a los casos dramáticos (buenos o malos), así como a los que apoyan las propias opiniones. El médico que dice que una intervención es excelente basado únicamente en su experiencia suele haber evaluado la "excelencia" a partir de los casos de pacientes que volvieron a recibir su atención. Y omite, involuntariamente, a quienes no volvieron a su consulta por variadas razones, entre las cuales puede encontrarse, fácilmente, que no hubo mejoría y el paciente decidió cambiar de médico.

Dado que el clínico no lleva habitualmente un registro tan preciso como para saber quién volvió y quién no, y por qué razones, el recuerdo tiende a estar sesgado. Uno de los autores recuerda cómo, al posesionarse como director de un puesto de salud durante su año de medicatura rural, el colega saliente le expresó su opinión, bienintencionada, que mencionaba cómo la población de la localidad era muy sensible a los medicamentos y respondía muy bien a sus tratamientos, pues ninguno de sus pacientes volvía a su consulta. "¡Todos se mejoran!", decía con asombrosa ingenuidad.

También puede ocurrir lo contrario. El clínico, ante una experiencia traumática o fuerte, le da un marcado valor a una reacción adversa poco común, y generaliza dicho riesgo. Quien haya tenido la mala experiencia de enfrentar un efecto adverso grave de un medicamento, aun si este es de verdad insólito, probablemente exagerará el riesgo de que tal hecho vuelva a ocurrir.

Libros de texto y fuentes bibliográficas similares

Aunque suelen contener gran cantidad de información relevante para la pregunta en cuestión, los libros de texto son, a menudo, la expresión escrita y ordenada de las opiniones de los autores (que son autoridades reconocidas en los temas). La evidencia que sustenta dichas opiniones no siempre está citada en el texto, y tampoco es fácil asegurar que el autor consultó evidencia tanto favorable como desfavorable a su opinión. De hecho, muy rara vez un libro de texto documenta las estrategias empleadas por el autor para identificar y evaluar las evidencias que sustentan sus opiniones.

Adicionalmente, al leer el libro de texto estamos recibiendo la opinión del autor acerca de la validez y los contenidos de las fuentes de información o de evidencia consultados por él, sin que se tenga la posibilidad de juzgar la validez de lo que él presente como evidencia. Finalmente, el proceso de redacción y edición de un texto toma meses o años. Generalmente, cuando un libro ve la luz, las referencias bibliográficas están desactualizadas; en ocasiones, hace varios años. Así, mientras las primeras publicaciones sólidas y concluyentes sobre la utilidad de aplicación de esteroides en casos de amenaza de parto pretérmino para disminuir la morbilidad y la mortalidad perinatales aparecieron entre 1972 y 1980, la inclusión de esta terapia como recomendación en los libros de texto de perinatología tardó ***diez años más***.

La literatura científica

En principio, la literatura científica es la fuente ideal de información, pues debe conducir al investigador a encontrar evidencia. Ahora bien, una buena parte de lo que se publica en la literatura corresponde a opiniones que suelen provenir de gente conocedora del tema, como es el caso de los artículos editoriales, las cartas al editor, las revisiones no sistemáticas de la literatura, las declaraciones y los consensos.

Suelen ser fuentes de evidencia, también, los artículos que informan los resultados de investigación: los llamados ***artículos originales***. **Estos se clasifican en dos grandes categorías:** 1) ***informes primarios de investigación*** y 2) ***artículos agregativos***.

Ejemplos de los primeros son los reportes de experimentos clínicos, los estudios de cohortes, los estudios de casos y controles, etc. Ejemplos de los segundos son las revisiones sistemáticas de la literatura (**RSL**) **—con o sin** metaanálisis—, las **GPC basadas en evidencia** y los análisis de decisiones.

Hasta hace relativamente poco tiempo, tener acceso a dichas fuentes de evidencia solía ser un proceso dispendioso y complejo, que requería el soporte permanente de expertos en búsqueda de información, centralizados en bibliotecas o en centros especializados. Afortunadamente, este proceso se ha simplificado, y hay en la actualidad múltiples herramientas en Internet que facilitan la búsqueda y la identificación de literatura relevante para la pregunta de investigación que se quiere responder, así como fuentes para obtener tal información en texto completo. A continuación se presentan las más destacadas.

Herramientas para la búsqueda de literatura científica

Las herramientas de búsqueda nos ofrecen un punto de acceso único a múltiples fuentes de información. La interfaz de esas herramientas nos permite ingresar los términos que seleccionamos para hacer la búsqueda de una manera amigable y funcional. Estos términos suelen ser las palabras clave que se han definido como las más relevantes para responder nuestra pregunta; sin

embargo, también se pueden ingresar búsquedas específicas, como, por ejemplo: autor, revista o título de un artículo.

Las herramientas de búsqueda presentan un listado de citaciones relevantes con múltiples opciones para filtrar o reordenar los resultados. Adicionalmente, cada citación nos ofrece un enlace para acceder al artículo en texto completo, a través de alguna de las herramientas diseñadas específicamente para tal fin. Las principales herramientas de búsqueda de información incluyen herramientas específicas para literatura biomédica, incluyendo PubMed y EMBASE, así como herramientas genéricas de múltiples disciplinas, como Google Scholar, Web of Science o Scopus[4]. Cada una de ellas tiene ventajas y desventajas.

PubMed

PubMed es la herramienta de búsqueda en línea más ampliamente utilizada, y se dedica a literatura biomédica. Fue desarrollada por el Centro Nacional de Información en Biotecnología (NCBI, por las siglas en inglés de National Center for Biotechnology Information), que hace parte de los Institutos Nacionales de Salud de Estados Unidos, y fue puesta en funcionamiento en 1996.

PubMed ofrece acceso gratuito a la más importante base de datos biomédica (Medline), además de otras bases de datos, como OLDMEDLINE (que incluye artículos publicados entre 1920 y 1967), PubMed CENTRAL (fuente de artículos en texto completo) y la base de datos de libros de la NCBI[5]. A la fecha, PubMed tiene indizados más de 24 000 000 de referencias, de las cuales más de 21 000 000 corresponden a artículos publicados en MEDLINE, incluyendo alrededor de 5600 revistas científicas del mundo entero, en más de 40 idiomas, que, a su vez, representan a 80 países. La base de datos se actualiza de forma permanente; incluye entre 2000 y 4000 referencias cada día.

Cabe aclarar que PubMed no es la única forma de acceder a la base de datos de MEDLINE: otros programas comerciales, como OVID o Proquest, también permiten el acceso a esta base, e incluyen funcionalidades diferentes.

Vale la pena resaltar algunas características de PubMed que facilitan la búsqueda de literatura biomédica[6,7]:

- Se pueden hacer las búsquedas ya sea en términos corrientes y cotidianos, o bien, en términos estandarizados (MeSH). Utilizar estos últimos hará más específica y completa la búsqueda.
- La página de resultados ofrece herramientas y opciones para limitar los resultados o modificar la búsqueda. Los principales filtros incluyen fechas de publicación, tipo de artículo, idioma, grupo etario, etc. Esta herramienta permite refinar la búsqueda evitando revisar de forma innecesaria grandes volúmenes de referencias. El uso de filtros permite una significativa mejora en la especificidad, pero tiene sus riesgos, pues, ocasionalmente, el filtro reduce de manera peligrosa el espectro. Con alguna frecuencia los usuarios sin mucha experiencia filtran por idioma, para evitar resultados en idiomas que no conocen.
- El error común está en que habitualmente se filtra para obtener artículos solo en inglés. Un poco más de 40 % de los artículos publicados provienen de Estados Unidos, pero del total de artículos de cualquier procedencia, el 93 % están publicados en inglés como idioma original. Ello parece dar a entender que se pierde poco cuando se filtra por inglés. Puede ser una buena solución para quien busca responder una pregunta clínica urgente, o para quien quiere revisar las generalidades de un tema. Pero para una revisión sistemática, para un metaanálisis o para algún otro tipo de actividad integrativa, es fundamental hacer la revisión más completa posible.
- Al filtrar por inglés, se omitirán los artículos en los demás idiomas, entre los cuales (lo olvidan a veces los usuarios) está… *¡el español!* Así pues, una búsqueda filtrada por idioma inglés omitirá la literatura procedente de Iberoamérica. Debe recordarse, además, que, aproximadamente, el 84 % del total de artículos tienen un resumen, escrito por los autores, en inglés, lo cual le permitirá al lector decidir si un artículo en ruso, en húngaro o en sueco es de interés para su búsqueda. Y si es así, deberá obtener la traducción del artículo.
- Permite la combinación de dos o más búsquedas a través de los operadores lógicos AND, OR y NOT ("y", "o" y "no", respectivamente). Estos operadores permiten enfocar la búsqueda y mejorar claramente la especificidad, sin sacrificio de la sensibilidad.
- Se presentan los detalles de la búsqueda hecha, y se muestra específicamente cómo

PubMed ha "traducido" los términos de búsqueda ingresados. Esta información facilita la reproducibilidad de la búsqueda y facilita su reporte en estudios integrativos, como revisiones sistemáticas de la literatura o GPC.

- Ofrece acceso al artículo en texto completo, si este se halla incluido en la base PubMed Central (La cual incluye más de 2,8 millones de artículos; o sea, alrededor del 10 % del total de las referencias incluidas en MEDLINE). O, al menos, presenta un enlace que lleva al usuario a fuentes pagas de texto completo. Dentro de la biblioteca de una institución, como un hospital o una universidad, cabe esperar que se haya hecho el trabajo de integración, para que el enlace lleve al usuario al texto completo en alguna de las bases de datos que suscribe la institución.
- Se puede limitar la búsqueda específicamente a artículos que se refieran a estudios clínicos (sin incluir ciencias básicas), o seleccionar únicamente RSL, o estudios de cohorte, o experimentos clínicos en cualquiera de sus fases clínicas, etc. Esta opción permite enfocar aún más la búsqueda y ahorra mucho tiempo y recursos.
- Se pueden seleccionar las referencias de interés y enviarlas, ya sea a un archivo electrónico o a un manejador de referencias bibliográficas, como EndNote o Reference Manager, lo que facilitará la escritura de los resultados de la investigación, o bien, el almacenamiento estructurado y la posterior consulta.
- En los resultados de la búsqueda, y en relación con cada referencia, PubMed identifica los artículos más relacionados con el tema tratado, teniendo en cuenta las palabras en el título, en el resumen y en los términos MeSH de cada registro. Esta herramienta facilita la búsqueda cuando hay una gran cantidad de literatura disponible.

Se puede acceder a PubMed a través de la página electrónica http://www.ncbi.nlm.nih.gov/pubmed, en la que, además de las herramientas para búsqueda, hay tutoriales de gran sencillez (de duración de unos pocos minutos, en lenguaje sencillo y comprensible, en inglés con subtítulos en inglés). No sobra recordar que la experiencia en búsquedas es uno de los pilares para la mejoría de la especificidad[2].

EMBASE

EMBASE fue desarrollada inicialmente en 1947, como *Excepta médica database*. Desde 2000, es manejada por Elsevier, un proveedor mundial de soluciones de información científica con énfasis en salud, y se ha convertido en una poderosa herramienta de búsqueda por suscripción. La principal ventaja de EMBASE radica en que incluye un número mayor de revistas que MEDLINE, lo que significa, también, una mayor representación geográfica. Se calcula que hay más de 6 000 000 de artículos y casi 3000 revistas adicionales. En la actualidad, tiene más de 28 000 000 de registros, provenientes de más de 8400 revistas médicas. Al igual que PubMed, la base de datos es actualizada a diario, e incorpora más de 900 000 nuevos títulos cada año. Este amplio cubrimiento la convierte en una de las bases de datos favoritas a la hora de hacer RSL.

Otra ventaja de EMBASE radica en el excelente cubrimiento en áreas de farmacología, incluyendo términos de búsqueda estandarizados (EMTREE), adicionales a los hallados en MEDLINE, para medicamentos y sus efectos adversos, y ello la hace ideal para temas de farmacovigilancia. Los medicamentos pueden consultarse por nombre químico, por nombre comercial o por el del fabricante; además, tiene un excelente cubrimiento en temas de biotecnología y equipos médicos.

Otras herramientas específicas para literatura biomédica

Vale la pena resaltar otras bases de datos, específicas de literatura biomédica, que se destacan en algunos campos específicos.

- La Biblioteca Cochrane, publicada desde 1997, es un producto de la Colaboración Cochrane, una organización internacional que prepara, actualiza y publica RSL biomédicas de alta calidad. Producto del interés de Archie Cochrane (pronúnciese Cókran, /'kɔ:xrən/), un médico escocés que promovió de manera incesante la importancia de los ECA, se inició como un centro de registro de experimentos clínicos, de sus resultados y de metaanálisis. Gracias a un acuerdo con BIREME (Centro especializado de la Organización Panamericana de la Saluda [OPS] radicado en Brasil), se puede tener acceso gratuito a la colaboración Cochrane estando en América Latina. La dirección de la pági-

na principal de la Colaboración Cochrane en Internet es: www.cochrane.org y la de BIREME es http://cochrane.bvsalud.org. Además de la destacada base de datos de RSL, la biblioteca Cochrane incluye un recurso llamado CENTRAL (http://community.cochrane.org/editorial-and-publishing-policy-resource/cochrane-central-register-controlled-trials-central), una lista bibliográfica de publicaciones sobre experimentos clínicos. Estos artículos son identificados por colaboradores de la Colaboración Cochrane que se esfuerzan por revisar la literatura científica mundial y crean una fuente de datos para las revisiones sistemáticas. Vale la pena, adicionalmente, destacar DARE, una base de datos con resúmenes (http://community.cochrane.org/editorial-and-publishing-policy-resource/database-abstracts-reviews-effects-dare) estructurados que valoran y sintetizan revisiones sistemáticas publicadas anteriormente y que se consideran de buena calidad.

- NICE (National Institute for Health and Care Excellence) http://www.evidence.nhs.uk/. Un organismo público del Ministerio de Salud del Reino Unido, cuya misión es producir guías basadas en evidencia para el apoyo a profesionales de la salud y a tomadores de decisiones; también, para desarrollar estándares de calidad y sistemas de evaluación para los servicios de atención social, salud pública y políticas de salud, y, finalmente, para ofrecer información a profesionales de la salud, para decisores y para administradores en salud. NICE ofrece recomendaciones basadas en evidencia en un número significativo de temas médicos, cuyo objetivo principal es promover la atención integrada en salud. También ofrece apreciaciones críticas de la efectividad clínica y económica de tecnologías en salud sobre productos farmacéuticos y biológicos, y también, sobre procedimientos, dispositivos y métodos diagnósticos.
- Trip (www.tripdatabase.com) es una herramienta de búsqueda clínica cuyo objetivo es brindar información proveniente de investigación de alta calidad, para el apoyo de clínicos y profesionales de salud. Se inició en 1997, como un recurso para clínicos, no solo mediante su colección de imágenes, videos, información educativa para profesionales y para pacientes, sino a través de un sistema de suscripción para actualización en los temas seleccionados por el usuario.
- CINAHL (Current Nursing and Allied Health Literature) es una herramienta de búsqueda manejada por EBSCO que se especializa en enfermería y ciencias afines de la salud (nutrición, fonoaudiología). En la actualidad, incluye más de 4,3 millones de referencias, provenientes, a su vez, de más de 5000 revistas científicas. http://www.ebscohost.com/nursing/products/cinahl-databases/cinahl-complete
- PsycINFO es una herramienta de la Asociación Americana de Psicología (APA, por las siglas en inglés de American Pshychological Association) especializada en búsqueda de literatura relacionada con ciencias del comportamiento. En la actualidad, incluye más de 3,7 millones de registros y 2500 revistas científicas. http://www.apa.org/pubs/databases/psycinfo/index.aspx
- Finalmente, vale la pena mencionar múltiples herramientas de búsqueda disponibles de manera gratuita en la red que utilizan MEDLINE y presentan la información organizada de forma especial para facilitar las búsquedas[8]. Algunas de ellas entregan los resultados ordenados por relevancia, o las asocian a conceptos específicos propuestos por el usuario (por ejemplo, RefMed, Quertle, Hakia, PubFocus).

Herramientas de búsqueda genéricas para múltiples disciplinas

A lo largo de los últimos años, han venido ganando espacio herramientas de búsqueda que incluyen información de múltiples disciplinas, además de las ciencias biomédicas. Dentro de estas herramientas, se destacan Google Scholar y otras herramientas por suscripción, como Web of Science y Scopus.

- Google Scholar (https://scholar.google.com/) fue lanzada en 2004 por Google. Teóricamente, esta herramienta de búsqueda tiene acceso a todos los artículos disponibles electrónicamente, e incluye tanto revistas revisadas por pares como las que no lo son. La herramienta permite acceso a fuentes de texto completo; adicionalmente, muestra los resultados ordenados por relevancia y presenta información del impacto académico de cada estudio[9]. Tiene algunas limitaciones, puesto que no es posible guardar

y replicar las búsquedas, dado que puede haber cambios con cada actualización, y no soporta programas de manejo de referencias bibliográficas.

- Scopus (http://www.scopus.com/) fue lanzado en 2004 por Elsevier. Se trata de una base (la más grande existente) de resúmenes y citaciones de la literatura revisada por pares, e incluye revistas científicas, libros y resúmenes de conferencias. Ofrece información en ciencias de la salud, así como sobre ciencias sociales, humanidades, artes y ciencias naturales. Incluye referencias de más de 22 000 revistas científicas, que incluyen, en total, más de 55 000 000 de registros.
- Web of Science (http://wokinfo.com/) es una herramienta por suscripción desarrollada por Thompson Reuters en 2004. Cubre más de 12 000 revistas científicas de diferentes áreas del conocimiento (ciencias sociales, ciencias naturales, artes y humanidades). Sirve para identificar patrones de búsqueda, sugiere términos para mejorar las búsquedas y filtra los resultados de acuerdo con las áreas de interés. Se destaca por sus estrictos criterios de selección para la escogencia de las revistas que incluye.

Fuentes de artículos en texto completo

La búsqueda en las distintas herramientas previamente mencionadas proporcionará, como se ha visto, la referencia bibliográfica exacta y el resumen del artículo; en muchas ocasiones, propondrá fuentes de texto completo, pero solo en un porcentaje limitado de los casos ofrece la opción de acceso gratuito.

En años recientes, se ha propuesto, cada vez más, retirar las limitaciones de costo para obtener acceso a la literatura científica a través de publicaciones con acceso abierto (***open access***). Se destaca dentro de tal iniciativa PubMed Central, que a la fecha ofrece más de 3 000 000 de artículos, de forma gratuita. Cabe aclarar que se mantienen los derechos de autoría.

Otras fuentes de artículos gratuitos incluyen el British Medical Journal (disponible a través de la página electrónica www.bmj.com), y grupos editoriales, tales como BioMed Central www.biomedcentral.com/

La dirección de Free Medical Journals (www.freemedicaljournals.com) ofrece a la fecha más de 4300 revistas que ofrecen sus artículos en texto completo, así como un servicio de alerta que por correo electrónico informará al suscriptor sobre las tablas de contenido de las revistas de interés. Adicionalmente, tiene una sección de libros que ofrecen acceso gratuito al texto completo.

Dentro de las bases de datos por suscripción se destacan:

- Springer's SpringerLink (http://link.springer.com/), que incluye más de 7,7 millones de artículos provenientes de libros y de revistas científicas, incluidos los materiales suplementarios.
- ScienceDirect (http://www.sciencedirect.com/), de Elsevier, contiene más de 11 000 000 de textos completos, provenientes de revistas, libros, e, incluso, de materiales audiovisuales.
- Wiley Online Library (http://onlinelibrary.wiley.com/), desarrollada por Wiley-Blacwell, y con más de 4 000 000 de citaciones.

Bases de datos latinoamericanas

Dado que una gran proporción de la literatura científica latinoamericana no está disponible en las bases de datos previamente mencionadas, vale la pena resaltar algunas bases de datos que ofrecen acceso a dichos artículos[10]:

- LATINDEX (http://www.latindex.unam.mx/) es el sitio del llamado Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal. Cuenta con enlace a 6161 de las 23 216 revistas que tiene en su directorio, incluyendo revistas tanto biomédicas como de otras áreas del conocimiento. El catálogo Latindex, puesto en línea en marzo de 2002, cuenta con una colaboración regional donde las revistas son inicialmente seleccionadas y clasificadas por cada país socio, según criterios internacionales de calidad editorial. En el momento, es administrado por la Universidad Nacional Autónoma de México (UNAM).
- PERIODICA es un portal de búsquedas, también administrado por la UNAM, (http://periodica.unam.mx/F?func=find-b-0&local_base=per01), y que ofrece alrededor de 350 000 registros bibliográficos de artículos originales, informes técnicos, estudios de casos, estadísticas y otros documentos,

publicados en cerca de 1500 revistas de América Latina y del Caribe especializadas en ciencia y tecnología.

- Redalyc (http://www.redalyc.org/) es la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal. Esta base, liderada por la Universidad Autónoma del Estado de México (UAEM), tiene como objetivo difundir la actividad científica que se produce en Iberoamérica, y que se materializa, desde 2002, en una hemeroteca científica en línea y de libre acceso para ciencias sociales y humanas. En la actualidad, ofrece más de 350 000 artículos en texto completo, y da acceso a más de 917 revistas científicas.
- SciELO (http://scielo.org/php/index.php?lang=en) es la Scientific Electronic Library Online-biblioteca virtual para Latinoamérica, el Caribe, España y Portugal. Se desarrolla como red en dos campos: Colecciones regionales (SciELO Brasil, SciELO Chile, SciELO Cuba, SciELO Colombia, etc.), y áreas temáticas (SciELO Salud Pública). Esta iniciativa, que surgió en Brasil, ha tenido gran impacto en Hispanoamérica. Su nodo en Colombia se encuentra en la Universidad Nacional, con un consejo que tiene representantes de Colciencias, de la OPS, de la Universidad Nacional y de representantes de las revistas indizadas.
- Imbiomed (http://www.inbiomed.org/), o Índice Mexicano de Revistas Biomédicas Latinoamericano, es un portal que contiene, aproximadamente, 350 revistas de Latinoamérica (60 de ellas, de Colombia) y sus artículos en texto completo. La base registra cuántas veces se consulta cada artículo y si se lo imprime o se lo recomienda.
- Lilacs (http://lilacs.bvsalud.org/es): esta base de datos (cuyo nombre viene de Literatura Latinoamericana y del Caribe en Ciencias de la Salud [LILACS]) es un producto cooperativo del Sistema Latinoamericano y del Caribe de Información en Ciencias de la Salud, administrada por el sistema Bireme, en conjunto con la OPS y la Organización Mundial de la Salud (OMS). Sus publicaciones son de América Latina y del Caribe, e incluyen, desde 1982, artículos de revistas y capítulos de libros, libros completos y otros documentos relacionados con el área de la salud. En la actualidad, indiza más de 568 000 artículos, de los cuales 300 000 están en texto completo.

Búsqueda de literatura gris

Cuando un investigador quiere hacer una RSL o quiere estar completamente actualizado en un tema en particular, debe hacer búsquedas y considerar la *literatura gris*, que se refiere a la que no ha sido publicada, o fue publicada por editoriales no comerciales.

El crecimiento de Internet permite ahora tener acceso a distintas fuentes para identificar este tipo de literatura[11]. Si se quiere identificar documentos como evaluaciones de tecnologías en salud, documentos gubernamentales y publicaciones de las casas farmacéuticas, la Academia de Medicina de Nueva York opera el sitio "Grey Literature Report" (http://www.greylit.org/), que ofrece acceso a literatura gris nueva en áreas de investigación médica y salud pública. Así mismo, la Agencia Canadiense para los Medicamentos y Tecnologías en Salud ofrece esta posibilidad a través de la página electrónica (http://www.cadth.ca/en/resources/finding-evidence-is/grey-matters). Otros portales electrónicos con la misma finalidad incluyen http://mednar.com/ y http://www.greynet.org/greysourcei ndex.html

Una buena fuente para identificar experimentos clínicos que fueron desarrollados recientemente y que aún no han sido publicados, o para ampliar la información sobre experimentos ya identificados, es: http://clinicaltrials.gov/, operada por la Biblioteca Nacional de Medicina de Estados Unidos. Una herramienta similar es la Plataforma de Registros Internacionales de Estudios Clínicos, desarrollada por la OMS[12] y disponible a través de http://www.who.int/ictrp/es/. Una opción alternativa es la Base Europea de Estudios Clínicos (https://eudract.ema.europa.eu/).

Referencias

1. Shariff SZ, Bejaimal SA, Sontrop JM, et al. Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. J Med Internet Res. 2013;15:e164.
2. Ilic D, Tepper K, Misso M. Teaching evidence-based medicine literature searching skills to medical students during the clinical years: a randomized controlled trial. J Med Libr Assoc JMLA. 2012;100:190-6.
3. Valcke M, De Wever B. Information and communication technologies in higher education: evidence-based practices in medical education. Med Teach. 2006;28:40-8.
4. Khare R, Leaman R, Lu Z. Accessing biomedical literature in the current information landscape. En: Kumar VD, Tipney HJ, editores. Biomedical Literature Mining. Nueva York, NY: Springer; 2014. p. 11-31.
5. Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 1. background and basic searching. J Telemed Telecare. 2013;19:479-86.
6. Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 2. display settings, complex search queries and topic searching. J Telemed Telecare. 2014;20:44-55.
7. Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 3. advanced searching, MeSH and My NCBI. J Telemed Telecare. 2014;20:102-12.
8. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database J Biol Databases Curation. 2011;2011:baq036.
9. Bramer WM, Giustini D, Kramer BM, et al. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev. 2013;2:115.
10. Gómez C. Indización, bases de datos y difusión de universitas médica. Univ Medica. 2007;48:187-92.
11. Citrome L. Communication of clinical trial results beyond PubMed: Searching the "Grey Literature" for Clinical Trial Results. Innov Clin Neurosci. 2014;11:42-6.
12. Lefebvre C, Glanville J, Wieland LS, et al. Methodological developments in searching for studies for systematic reviews: past, present and future? Syst Rev. 2013;2:78.



SECCIÓN

2

Muestreo, medición y variabilidad

CAPÍTULO 4 Variables y fuentes de error en la medición. Confusión, sesgo, azar
Juan Gabriel Ruiz Peláez

CAPÍTULO 5 Población y muestra
Juan Gabriel Ruiz Peláez, María Nelcy Rodríguez Malagón

CAPÍTULO 6 Cálculo del tamaño de muestra
Martín Alonso Rondón Sepúlveda

CAPÍTULO 7 Variables no observables en epidemiología clínica: el uso de constructos
Socorro Moreno Luna

CAPÍTULO 8 Adaptación cultural y validación de escalas
Carlos Gómez-Restrepo, Martín Alonso Rondón Sepúlveda, María Beatriz Ospina García

CAPÍTULO 9 Medidas de frecuencia, de asociación y de impacto
Juan Manuel Lozano León

CAPÍTULO 10 Uso de pruebas diagnósticas en medicina clínica
Álvaro J. Ruiz Morales

CAPÍTULO

4

Variables y fuentes de error en la medición. Confusión, sesgo, azar

Juan Gabriel Ruiz Peláez

INTRODUCCIÓN

La representación mental que nos hacemos de los mundos que nos rodean y el recuerdo que tenemos y consultamos de estas representaciones constituyen la esencia de lo que se suele denominar ***conocimiento***. Este capítulo busca explorar y entender uno de los procesos centrales de la construcción de ***conocimiento científico***, que es la percepción sistemática y ordenada (medición) de características de la realidad. Se revisarán los conceptos de variable, enfoque y nivel de medición, papel analítico de las variables, error en la medición y en sus fuentes y los tipos de error que pueden ocurrir: ***sistemático*** y ***aleatorio***.

Para entender y sistematizar el proceso de medición de variables, es necesario adoptar algún tipo de enfoque epistemológico. En el presente capítulo se usa el enfoque propuesto por el filósofo de la ciencia Karl Popper, quien afirma que la consciencia percibe cuanto ocurre en tres ***mundos*** o *planos de realidad*, que tienen existencia objetiva y son susceptibles de ser explorados y conocidos. El mundo uno, o físico; el mundo dos, o de las experiencias subjetivas de los observadores (incluye pensamientos y emociones), y el mundo tres, o de las ideas y la cultura. Los fenómenos y los entes que "habitan" dichos mundos son percibidos y "medidos" por la consciencia a través de la percepción de sus características.



Más allá de Popper: Los "tres planos" de Penrose

En su libro *El camino a la realidad*, Roger Penrose, físico matemático, propone un enfoque epistemológico basado en tres "planos" o "realidades", que le permite explicar su visión de la relación entre las matemáticas, las "leyes" científicas (físicas) y la realidad observable. Al parecer, Penrose llega de forma independiente a esta proposición, pero sus planteamientos tienen una gran convergencia con los de Popper. El primer "plano" corresponde al universo real, material; es un marco espacio-temporal donde hay energía y materia interactuando a lo largo del tiempo, y en el cual surgen la vida y la consciencia. La consciencia, que surge del mundo físico, se desenvuelve en un segundo "plano": el de los sentimientos, los pensamientos, la memoria y la imaginación. Los seres vivos conscientes, los seres humanos, habitamos esta realidad física y nos hacemos una representación mental, ***imaginada***, de dicha realidad. A través de la razón, la consciencia accede a ese tercer plano, que, según Penrose, no es imaginado (creado), sino "descubierto" (tiene existencia autónoma, independientemente de que la consciencia lo examine o no), una especie de realidad virtual, pero con existencia objetiva. En esa realidad virtual a la que accede la consciencia, la razón permite conocer una serie de entes que tienen existencia por fuera del tiempo y el

espacio: abstracciones y representaciones de conceptos, los cuales siguen leyes estrictas que la razón puede descubrir. Los conceptos matemáticos y geométricos, entre otros, son los habitantes de ese universo virtual. El triángulo equilátero en un plano euclidiano existe como concepto y como entidad: sus lados son iguales, la suma de sus ángulos internos es 180º. El concepto, la abstracción "triángulo", tiene existencia independientemente de si hay una consciencia que la piense o la imagine.

Ahora bien, dichos entes abstractos que existen en el plano virtual (sobre todo, las leyes matemáticas) parecen subyacer al funcionamiento físico del universo (las "leyes") y explicarlo. Esta visión se corresponde parcialmente con el planteamiento de Karl Popper del ***mundo tres***.

Las cosas que existen en los tres mundos (o en los tres planos) son percibidas por la consciencia; las del mundo uno, o físico, a través de los órganos de los sentidos (o de extensiones de dichos órganos, como el microscopio, las pruebas químicas o biológicas, etc.); las del mundo dos, a través de la autoconsciencia, la capacidad interna de percibir y "ver" u "oír" nuestros pensamientos, nuestros recuerdos y nuestras emociones, y, finalmente, las del mundo tres, a través de la razón, e instrumentos asociados al uso de la razón (por ejemplo, fórmulas, ecuaciones, algoritmos, etc.).

Sin importar de cuál mundo o plano se trate, la consciencia no percibe de forma directa los entes, los eventos y los fenómenos que viven y ocurren en esos planos de existencia. La consciencia recibe información, que es interpretada como "características" provenientes de los hechos, los fenómenos y los entes percibidos. Por ejemplo, cuando la consciencia reconoce un ente en el mundo físico e interactúa con él, percibe características (color, forma, olor, movimiento, sonido, etc.) que ubica dentro de un determinado marco espacial y temporal. Algunas de esas características del ente percibido no cambian a lo largo del tiempo o cuando su posición en el espacio varía, pero muchas otras sí cambian. Todas las características que la consciencia pueda percibir y puedan cambiar de un ente a otro o en el mismo ente o fenómeno al cambiar el tiempo o el lugar se denominan *variables*. **Las variables de los planos uno y** tres suelen percibirse como "objetivas"; diferentes observadores, usando correctamente los mismos instrumentos, deben obtener las mismas mediciones. Las variables del plano dos son esencialmente subjetivas, relativas al sujeto que las experimenta, y no mensurables de forma directa por un observador externo. Estas variables suelen jugar un papel predominante en el campo de las ciencias sociales y de las metodologías cualitativas.

DEFINICIÓN DE VARIABLE

Sin importar cuál mundo habitan, distinguimos un ser o un evento de otro por el hecho de que percibimos características diferentes. Si de dos seres o eventos distintos percibimos exactamente las mismas características, nos sentimos incapaces de distinguirlos entre sí.

Una característica uniforme para diferentes elementos de un conjunto, o que no cambia en un mismo elemento a lo largo del tiempo, es una *constante*. **En contraste, cualquier característica** que podamos percibir (o medir) y que cambie de un sujeto a otro, o en el mismo sujeto a lo largo del tiempo, se denomina ***variable***.

En un bosque de pinos, el tronco de todos los árboles está compuesto, entre otras cosas, de madera. Esta característica es una *constante* **en los elementos de dicho conjunto.** En contraste, atributos como la altura y el número de ramas de cada árbol son ***variables*** que pueden diferir de los demás árboles, y también cambiarán en el mismo árbol, con el paso del tiempo.

En síntesis, el término ***variable*** se refiere a una característica de un ente, de un conjunto de entes, de un fenómeno o de conjunto de fenómenos o de eventos, y que puede adoptar diferentes valores (cuantitativo) o estados (cualitativo).

Hay variables cuyo valor numérico específico es indeterminado antes de ser medido, pero cuya frecuencia de ocurrencia está determinada por una función matemática definida —función de probabilidad— (***variables aleatorias***) y variables en las cuales el valor que adoptan se determina por mecanismos no probabilísticos (***variables no aleatorias***).

Como ejemplo de las primeras, podemos citar la edad de un paciente seleccionado aleatoriamente en la sala de espera de urgencias. Como ejemplo de las segundas, la longitud de la hipotenusa de un triángulo rectángulo, que está determinada por las de los catetos; si cambia la longitud de al menos uno de los catetos, también lo hace la de la hipotenusa, y solo hay un valor posible, el cual depende de unas relaciones precisas entre las variables, y no de una distribución de probabilidades (o de una distribución de probabilidades con cero grados de libertad). Recordemos que cuando el valor que una característica puede adoptar no varía a lo largo de una serie de circunstancias diferentes, se está hablando de una ***constante***. **Así, en las** mediciones relacionadas con la velocidad a la que caen los objetos por acción de la gravedad en la Tierra, la masa del objeto (***m***) **y el tiempo** transcurrido (***t***) **desde el inicio de la caída son** variables, mientras que la aceleración debida a la fuerza de gravedad (***g***) **es constante.**

CLASIFICACIÓN DE LAS VARIABLES

Las variables se pueden clasificar de acuerdo con varios criterios: a) según la naturaleza de la característica medida, b) según el enfoque de la medición, y c) según el nivel de medición empleado y su consecuencia, que es la forma de presentar (o resumir) el resultado de la medición. Adicionalmente, se puede describir el papel que representa una variable en describir o explicar una situación o un problema; es decir, su ***papel*** *analítico*.

Naturaleza de las variables

En primer lugar, la ***naturaleza*** de las variables puede ser ***discreta o continua.***

A pesar de que el hecho de que se asigne una cantidad numérica a la variable no es estrictamente parte de su naturaleza, sino del enfoque de medición (bien sea cuantitativo o cualitativo), hay tipos de variables discretas en las que puede ser natural asignar magnitudes. Por eso, resulta apropiado añadir un atributo más a la especificación de la naturaleza de las variables discretas, y es que, de forma natural, sean "numéricas" o "no numéricas" **(cuadro 4.1).**

Variables discretas

La naturaleza de una variable es ***discreta*** si se trata de una característica que solo puede variar a lo largo de estados o de clases definidos, con límites claros y mutuamente excluyentes. Ello implica que no existen estados "intermedios" entre estados consecutivos.

EJEMPLO

La variable ***sexo***, que nos permite clasificar a los seres humanos, solo cae en estados discretos: intersexual, masculino y femenino.

Cuando se enumeran los elementos de un conjunto que tienen una característica en común, se genera un tipo especial de variable discreta, cuya naturaleza es ***numérica***, o ***de*** *conteo*. **Ejemplos: Al medir aspectos como** muertes, infartos de miocardio, piezas dentarias con caries, etc., los eventos en sí son discretos: un sujeto está vivo o muerto, presenta o no un infarto de miocardio, o una pieza dentaria tiene o no caries; pero la frecuencia con la que ocurren es un fenómeno numérico: diez muertes, quince infartos, tres piezas con caries.

Debido a que el evento o el fenómeno en sí son discretos, no es posible que existan conteos intermedios entre eventos sucesivos; por ejemplo, el número de hijos por familia solo puede adoptar valores enteros, ***y no fraccionarios***, **aun** cuando al resumir información sobre conteos de eventos o de fenómenos se empleen a menudo números fraccionarios, los cuales, si bien ayudan a entender situaciones, no tienen representación en el mundo real: el número promedio de hijos

CUADRO 4.1 Variables clasificadas según su naturaleza

- Discretas
  - Numéricas (conteo)
  - No numéricas (categóricas)
    - **Dicótomas**
    - **Politomas (más de dos categorías)**
- Continuas (numéricas)

por familia en una comunidad determinada es de 2,66; pero ninguna familia de dicha comunidad puede literalmente tener 2,66 hijos.

Con la excepción ya mencionada de las variables de conteo, las variables discretas tienen intrínsecamente una naturaleza *no numérica*; es decir, no se les puede asociar de forma natural un número que represente una magnitud.

Algunas de las variables discretas no numéricas, dada su naturaleza, solo pueden adoptar uno de dos estados, mutuamente excluyentes. Por ejemplo, el estado vital de un ente solo puede ser *vivo* o *inanimado*, o un interruptor eléctrico solo puede estar encendido o apagado. Estas variables discretas, que, dada su naturaleza, solo pueden caer en uno de dos estados opuestos y mutuamente excluyentes, reciben varias denominaciones: *binarias*, *lógicas* (verdadero-falso) o, más comúnmente, *dicótomas* (del griego *diká*, o dos, y *tomé*, o cortado, dividido).

En los demás casos, las variables discretas no numéricas pueden, de forma natural, corresponder a más de dos estados mutuamente excluyentes. En contraste con las anteriores, se denominan discretas, no numéricas, *politomas* (del griego *poly*, o varios, y *tomé*, o dividido). Por ejemplo, si solo hay dos pigmentos disponibles, como blanco y negro, para colorear un objeto, la variable "color del objeto" solo puede adoptar uno de tres estados mutuamente excluyentes: blanco, negro o gris (no importan los matices de gris, todos tienen en común que no son absolutamente blancos ni absolutamente negros).

Variables continuas

Las variables continuas pueden adoptar infinito número de valores a lo largo de un continuo.

La exactitud con la que se mide la variable depende de la sensibilidad analítica (resolución o aproximación) del proceso de medición. La naturaleza de la totalidad de las variables continuas es no solamente continua, sino numérica: a la variable no se le asigna una magnitud: la variable continua en sí misma *es* una magnitud, vectorial o escalar.

EJEMPLO

Las magnitudes físicas, como la masa, la velocidad, la temperatura o la longitud, son ejemplos de variables continuas. El peso de una persona puede variar a lo largo de un número infinito de valores posibles. La cantidad de valores intermedios entre dos puntos de la medición (por ejemplo, 50 kg y 51 kg) es ilimitada, y depende de la aproximación o del detalle (sensibilidad analítica) con los que se pueda efectuar la medición.

Enfoque de medición

De acuerdo con el *enfoque de medición*, o la *forma de ser medidas*, la medición de una variable puede ser cuantitativa o cualitativa. Es *cuantitativa* si se asigna un valor numérico cardinal (discreto —números enteros— o fraccionario —números reales—) a la magnitud de la variable.

Es *cualitativa* si los valores de la variable se clasifican en categorías que describen una cualidad o un estado discretos, y no una cantidad **(cuadro 4.2)**.

Enfoque cualitativo

Las variables de naturaleza discreta suelen medirse en términos cualitativos.

El atributo, o característica observada, es clasificado de acuerdo con la presencia o la ausencia de un estado particular en una de dos o más categorías o clases definidas, y mutuamente excluyentes. En el enfoque cualitativo, el observador que hace la clasificación compara el estado del atributo medido con una definición, o referente (reconocimiento de un patrón). El elemento cuyo atributo es evaluado es, entonces, clasificado en la categoría o el estado para el que satisfaga la definición.

CUADRO 4.2 Enfoques de medición de variables

- Cualitativa (calificación)
  - No numéricas (discretas)
  - Numéricas categorizadas
- Cuantitativa (medición)
  - Numéricas
    - ✧ **Discretas**
    - ✧ **Continuas**

Si bien en el enfoque cualitativo no hay una cuantificación del atributo, la descripción del estado puede incorporar la intensidad de la característica de interés.

EJEMPLO

Al clasificar el atributo "comodidad" de una silla con una perspectiva cualitativa, el observador puede limitarse a clasificarla como "cómoda" o "incómoda"; o puede clasificarla en varias categorías ordenadas según la "intensidad", o grado de comodidad. No hay una magnitud asociada al estado específico de comodidad, pero sí hay un ordenamiento que refleja su comodidad frente a otras sillas. La incorporación o no de la intensidad absoluta o relativa del atributo evaluado cualitativamente se refleja, como se discutirá más adelante en este mismo capítulo, en el nivel o la escala de medición que se haya empleado.

Como ya se mencionó, las variables de naturaleza continua se asocian de forma clara a valores numéricos, y suelen ser medidas con un enfoque cuantitativo. Las variables numéricas continuas pueden, sin embargo, ser medidas en términos no cuantitativos, al ser transformadas en miembros de clases o de categorías. Este proceso se denomina ***categorización.***

EJEMPLO

La evaluación de la presión arterial se hace, principalmente, a través de las cifras de presión sistólica y diastólica en milímetros de mercurio (mm Hg), que es una variable numérica continua. Sin embargo, para facilitar la aproximación terapéutica y pronóstica, se puede establecer una categorización de la presión arterial, que tiene gran utilidad práctica.
La clasificación propuesta en 2003 por el VII Comité Nacional Conjunto para Diagnóstico y Tratamiento de la Hipertensión arterial (VII Joint National Committee) ofrece las categorías *Normal* para una presión sistólica < 120 y diastólica < 80; *Prehipertensión*, para sistólicas entre 120 y 139 o diastólicas entre 80 y 89; *Hipertensión Estado 1*, para presiones sistólicas entre 140 y 159 o diastólicas entre 90 y 99; *Hipertensión Estado 2*, cuando la presión sistólica sea > 160, o la diastólica, > 100[1].

Enfoque cuantitativo

Las variables continuas son, fundamentalmente, magnitudes; o sea, cantidades numéricas de una característica.

El enfoque que naturalmente encaja con estas variables es el cuantitativo. En lo que respecta a las variables discretas, hay, cuando menos, dos circunstancias en las cuales se utiliza un enfoque cuantitativo:

Cuando el atributo o el fenómeno medido no es una característica de un solo elemento, sino la frecuencia de ocurrencia del atributo en un conjunto de elementos, o la frecuencia de un evento. Como ya se mencionó, esta circunstancia se refiere a las variables de naturaleza numéricas discreta o variables de conteo.

Cuando se intenta cuantificar el grado o la intensidad de una cualidad en términos numéricos (determinar la magnitud de la intensidad). Por ejemplo, la presencia o la ausencia de dolor puede graduarse en categorías discretas (enfoque cualitativo, escala de medición ordinal, según como se verá más adelante), o puede pedírsele al paciente que lo cuantifique como una cantidad a lo largo de un continuo (escala visual análoga).

Nivel o escala de medición

En el ámbito operativo, la medición de la variable se lleva acabo utilizando una escala, o sistema de medición **(cuadro 4.3)**. El cuadro muestra las escalas de acuerdo con el nivel de información que estas recolectan de la variable, desde el más elemental (nominal, nombre) hasta los más complejos.

Escalas de medición apropiadas para variables discretas

Los ***niveles de medición***, o *escalas de medición* apropiadas para las variables discretas no numéricas son: a) nominal, b) ordinal, o de rango, y c) de pertenencia a categorías ordenadas (caso especial de las escalas ordinales). El número y la calidad de operaciones que se

CUADRO 4.3 Niveles o escalas de medición

- Nominal
- Ordinal
  - Pertenencia graduada a categoría
  - Rango (ordinal propiamente dicha)
- Intervalo
- Razón
- Otras (log-intervalo, razón derivada, etc.)

pueden hacer con las mediciones nominales y ordinales se limitan a resumir la información mediante conteos de frecuencias absolutas y relativas (proporciones).

Escala o nivel nominal

Al medir la característica "marca" de un automóvil, el resultado se expresa en términos del nombre de la marca: Mercedes Benz, Peugeot, Chevrolet, etc. Esta escala de medición (enfoque cualitativo) de características discretas se llama ***nominal.***

EJEMPLO

Al revisar los diagnósticos de los pacientes en una unidad de cuidado intensivo (uci), se encuentra que hay casos de infarto del miocardio, de tromboembolismo pulmonar, de sepsis de origen abdominal, de síndrome de Guillain-Barré y de hemorragia subaracnoidea. Dichos diagnósticos no tienen un orden intrínseco, y no pueden ordenarse excepto cuando se especifique un criterio externo, tal como órgano afectado, severidad, necesidad de ventilación mecánica o de monitorización invasiva.

El uso de una escala nominal implica asignar un nombre al atributo o estado medido, que denota la pertenencia a una categoría o clase. El "valor" que adopta la variable es solo una etiqueta, que puede ser un número o cualquier otro símbolo. Por ejemplo: Sexo Masculino (código 1); Sexo Femenino (código 2). Independientemente del nombre o el código asignado, la clasificación en una categoría no implica un ordenamiento o una jerarquía del atributo medido; por ejemplo: religión, raza o el número en la camiseta del jugador de un equipo de fútbol.



Una escala nominal se aplica correctamente a variables cuyos estados, de forma natural, no representan un ordenamiento de acuerdo con la magnitud de alguna otra característica, y, por tanto, no tienen ninguna implicación jerárquica.

Por ejemplo, sujetos clasificados según su grupo sanguíneo en las categorías **A**, **B**, **AB** y **O.**

Escala ordinal



Si se propone una clasificación para una variable discreta, que implique un orden intrínseco, la escala ya no es simplemente nominal, pues ahora, sin especificar criterios extrínsecos, los resultados de las mediciones tienen una jerarquía y pueden ser ordenados. El nivel de medición ordinal se aplica a variables de naturaleza discreta, o a continuas categorizadas.

Este nivel de medición tiene dos matices: 1) membresía graduada y 2) escala de posición o rango (ordinal propiamente dicha).



En el caso de la *membresía graduada* (pertenencia a categorías ordenadas), los sujetos son clasificados en una categoría de acuerdo con la intensidad del estado o atributo, sin tener en cuenta los resultados de la misma medición en otros sujetos del conjunto evaluado.

Las categorías pueden ser, por ejemplo, el grado de severidad del edema de los miembros inferiores, clasificado como ***leve*** (aumento del volumen de la extremidad, sin fóvea), ***moderado*** (hay fóvea del tobillo hacia abajo) o ***severo*** (fóvea por debajo de la rodilla). El edema de cada sujeto de un conjunto es comparado con las definiciones de los tres grados de edema, y clasificado según la que le corresponda, sin tener en cuenta si su edema es más o menos severo que el de los demás pacientes medidos. En síntesis, se reconoce que un nivel de medición es de membresía graduada cuando: a) la evaluación de la variable clasifica a los sujetos en categorías discretas, b) las categorías tienen ordenamiento o jerarquía (a diferencia de las escalas nominales) y c) las distancias entre categorías no necesariamente son simétricas.

Si bien el enfoque de medición es cualitativo, las categorías se ordenan de acuerdo con la intensidad del atributo medido, y debido a ello, algunos autores denominan a este enfoque "semicuantitativo". Tal denominación se usa, sobre todo, cuando para clasificar al sujeto se usan simultáneamente criterios cuantitativos y cualitativos.

EJEMPLO

En la evaluación ordinal de la bacteriuria en el examen del sedimento urinario se usa un enfoque semicuantitativo (se cuentan bacterias, pero se miran aspectos no cuantitativos) para categorizar los resultados. Las distancias entre categorías no son simétricas; es decir, "++" no representa una unidad más de cantidad de bacterias que "+" **(cuadro 4.4)**.

La escala de medición *ordinal* propiamente dicha se utiliza cuando la característica medida no es una cualidad (leve, moderado y grave), sino el lugar relativo que ocupa un elemento en un conjunto.

Esta escala se puede denominar ***de posición*** o ***de rango***, **equivalente al *ranking* en inglés.** Así, por ejemplo, una característica medida de forma cualitativa, como la gravedad del edema de los miembros inferiores (leve, moderado y grave) clasifica a los sujetos como miembros de tres categorías ordenadas (jerárquicas). Aun si el nivel de edema se cuantifica en términos de aumento de peso, puede seguir categorizándose en dichas clases. Sin embargo, los sujetos medidos también pueden ordenarse según la intensidad del edema respecto a la de los demás miembros del conjunto de sujetos al que pertenece el paciente sujeto de interés. Así, será primero el que tiene mayor exceso de peso por edema; segundo, el que le sigue; tercero, el siguiente, y así sucesivamente. El tercer sujeto puede estar en la categoría de moderado, y su peso, ser muy cercano al segundo, y los dos pueden estar muy lejos del primero y del cuarto. Los números no reflejan magnitud ni gravedad, sino solo el orden relativo de los elementos. Otros ejemplos incluyen el orden de llegada en una carrera, la posición en las notas de un examen, etc. **(cuadro 4.5)**.

CUADRO 4.4 Categorización (semicuantitativa) de la bacteriuria en el sedimento urinario (se examinan 40 campos secos fuertes)

0	**No bacterias o escasas**
+	**1-10 bacterias por campo**
++	11-100 bacterias por campo
+++	> 100 bacterias por campo, discernibles individualmente
++++	> 100 bacterias por campo, aglomeradas

Las nociones de membresía graduada y de rango se pueden combinar en las clasificaciones de sujetos que utilizan una ***distribución percentilar***. **Para generar las categorías, o clases ordenadas,** se obtiene la distribución de una variable numérica de interés en una población de referencia. Los sujetos de la población de referencia son ordenados de acuerdo con la magnitud de la variable de interés, y su posición relativa respecto a los valores de los demás elementos de la población de referencia se expresan ***centilarmente***; es decir, en 100 clases o categorías ordenadas. Por ejemplo, se va a generar una distribución percentilar con los resultados de una prueba de conocimientos. Supongamos que hay 400 sujetos en la población de referencia. De esta manera, los 4 sujetos con la mayor puntuación en la prueba ocupan el lugar más alto (el puesto 100); los 4 siguientes, el 99; los 4 siguientes, el 98, y así sucesivamente. Supongamos que los 4 sujetos del puesto 95 obtuvieron notas entre 4,6 y 4,7 sobre 5. Un sujeto que no pertenece a la población de referencia y que obtiene 4,7 en su examen es miembro de la categoría "percentil 95".

Operaciones con escalas nominales y ordinales

El número y la calidad de las operaciones que se pueden hacer con las mediciones nominales y ordinales se limitan a resumir la información mediante conteos de frecuencias absolutas y relativas (proporciones). La única medida de tendencia central que se puede hacer en una medición nominal es calcular la moda.

En una variable medida ordinalmente, es apropiado hacer comparaciones como "mayor que" o "menor que" entre individuos medidos. Aun si se asignan códigos numéricos a las categorías o mediciones, estos números representan una etiqueta (nominal) o un grado (rango u ordinal), y no tiene sentido hacer operaciones ma-

CUADRO 4.5 Características de las escalas de medición ordinales propiamente dichas, o de rango

- Ordenamiento según una medición externa: orden de llegada
- Distancias asimétricas
- Ejemplos: rango de calificaciones
  - Rango 1. Mejor nota de examen (99 / 100)
  - Rango 2. Segunda mejor nota (75 / 100)
  - Rango 3. Tercera mejor nota (74 / 100)
- La distancia entre la primera y segunda mejor nota es diferente a la distancia entre la segunda y tercera

temáticas simples, como sumas o restas, ni tales operaciones se pueden resumir con promedios y medidas de dispersión relacionadas.

EJEMPLO

En la codificación del sexo de un grupo de pacientes se asignaron códigos numéricos: 1 para femenino, 2 para masculino y 3 para indeterminado. Si se quiere resumir la distribución por sexo de dicho grupo, es necesario reportar el número y la proporción de mujeres (1), de hombres (2) y de indeterminados (3). Reportar el "promedio" de sexo (sumar los conteos de sujetos (1), (2) y (3) dividido por el total de sujetos) no informa nada interpretable, y carece de sentido: "el sexo promedio del grupo fue 2,1". Tampoco permite comparar grupos.

EJEMPLO

También en el caso de mediciones ordinales, no tiene sentido hacer adiciones, restas o cocientes. Suponga que tiene a un grupo de pacientes con edema. Los clasifica, de acuerdo con la severidad del edema, en grados: 1 (leve) 2 (moderado) 3 (severo) y 4 (anasarca). Quiere, además, describir la población en términos de intensidad o severidad del edema, antes y después de ofrecer una dieta baja en sodio y de la administración de un diurético. La forma apropiada es informar los cambios en las proporciones de los cuatro grados de edema, antes y después del tratamiento. Cualquier otra operación sería inapropiada. Dentro del espectro de comparaciones erróneas están: 1) sustraer del nivel de edema pretratamiento el nivel de edema postratamiento, 2) "promediar" los niveles de edema pre y post y compararlos (diferencia o cociente), y 3) variantes de esas dos operaciones. El número asociado a cada nivel no representa la magnitud del edema (cantidad cuantificable numéricamente), sino su "intensidad" o su posición relativa (rango). La distancia en magnitud entre un edema grado 1 y 2 no es uniforme en todos los sujetos; además, es diferente de la distancia entre un edema 2 y uno 3.

En el caso de las variables discretas que son medidas cuantitativamente (conteo), la escala usada para las variables numéricas discretas es, simplemente, el conteo del número de elementos o eventos, y dicho conteo es la forma de presentar y resumir este tipo de variables. Se puede, según cada caso específico, considerar la conveniencia y racionalidad de resumir la información con medidas de tendencia central y de dispersión.

En especial, cuando se hacen comparaciones utilizando razones que involucren variables numéricas discretas, tal como ya se mencionó (Ejemplo: en promedio, hay 2,5 hijos por familia en una localidad).

Escalas de medición apropiadas para variables continuas

En el caso de las variables continuas medidas de forma cuantitativa, las dos escalas de medición más utilizadas son ***intervalo*** **y** ***razón*****.**

Escala de intervalo

Se trata de una medición cuantitativa (numérica) continua, donde el cero de la escala es arbitrario, escogido por conveniencia, y las distancias entre valores consecutivos son iguales y significan lo mismo.

Un ejemplo típico de esta escala lo constituye la temperatura medida en grados centígrados, o Celsius (°C). En la escala centígrada, la temperatura se mide en una escala de intervalo; el cero es arbitrario, y corresponde a la temperatura de licuefacción del hielo a nivel del mar, es decir, 273,2 grados por encima del cero absoluto, o Kelvin (273,2 K), y 100 °C es la temperatura de ebullición del agua a nivel del mar (373,2 K). Un grado centígrado (1 **°C) es 1/100 de la diferencia** de temperatura entre el punto de fusión del hielo y el de ebullición del agua, y todos los *intervalos* **entre valores representan lo mismo** a lo largo de toda la escala: la diferencia de temperatura de 1 ml de agua a una temperatura de 2 °C y otro a 3 °C es la misma que entre 1 ml 100 °C y otro a 101 **°C. De ahí el nombre de** *escala de intervalo*. **Los valores negativos tienen** existencia real: una temperatura de -20 **°C no es** "ausencia" o "deuda" de temperatura, sino que equivale a 253,2 K.

Escala de razón

Cuando se mide una variable en una escala de razón, el punto cero de la escala es real y absoluto (significa ausencia de la variable), y no existen valores negativos.

Tal como ocurre en la escala de intervalo, la distancia entre dos valores consecutivos es igual y significa lo mismo. La mayoría de los resultados de laboratorio clínico son variables de razón (glucemia, colesterol total sérico, sodio sérico), al igual que las características físicas, como la estatura y el peso. De la misma manera, la temperatura medida en grados Kelvin también tiene un valor de cero absoluto, equivalente a la ausencia total (teórica) de temperatura.

Operaciones con escalas de intervalo y de razón

Las magnitudes medidas en escalas de intervalo y de razón se expresan utilizando el conjunto de los números reales. En el caso específico de las variables de razón (cero absoluto), solo se usan números reales positivos. La información medida se puede resumir utilizando medidas de tendencia central y de dispersión apropiadas para la distribución subyacente: medias, medianas, modas y desviaciones estándar y rangos, según como sea apropiado.

Las variables medidas en escala de intervalo permiten hacer sumas y restas, pero debido a que el cero es arbitrario, ni las multiplicaciones, ni las divisiones, ni otras operaciones más complejas (potenciación, transformaciones logarítmicas, etc.) tienen sentido. Por el contrario, las variables medidas a nivel de razón permiten hacer todas las operaciones matemáticas.

Las principales características, similitudes y diferencias entre las escalas de intervalo y de razón se resumen en el **cuadro 4.6**.

Relaciones entre variables

Para la planeación y el desarrollo de procesos de investigación clínica o epidemiológica, resulta conveniente clasificar las variables por medir según el papel que desempeñan en el proceso de análisis e interpretación de los resultados.

En las investigaciones descriptivas de enfoque cuantitativo, la respuesta a la pregunta de investigación se obtiene mediante la medición de una o más características de los elementos que conforman un conjunto determinado. El proceso se centra en ***describir*** o ***caracterizar*** conjuntos de elementos (poblaciones) midiendo variables en cada sujeto (censo) o en muestras representativas de la población de interés, y resumiendo sus valores o sus resultados. Todas las variables medidas cumplen papeles analíticos similares: describir o caracterizar. En estas investigaciones descriptivas, de enfoque cuantitativo, las variables medidas no tienen ninguna jerarquía particular diferente de la que les otorga el hecho de ser la respuesta a la pregunta de investigación.

Por el contrario, en investigaciones analíticas (observacionales o experimentales) el objetivo principal del proceso investigativo es la evaluación de una hipótesis. Dichas hipótesis, usualmente, se formulan en términos de la presencia o la ausencia de asociaciones entre la exposición a uno o más factores y la aparición de uno o más desenlaces, mientras se controla el efecto de otras características o fenómenos que pueden o podrían confundir o modificar los efectos de las supuestas asociaciones, en caso de que existieran.

En estas circunstancias, las variables se pueden agrupar en tres grandes categorías: 1) variables ***de exposición***, 2) variables ***de desenlace*** y 3) variables ***de control***. Las variables de exposición, también denominadas ***independientes de interés***, o *predictoras*, son las que expresan la asignación a las diferentes exposiciones a eventos o factores cuyos efectos son investigados. Las variables de desenlace, o ***dependientes***, son las que se refieren a si se produce o no el efecto postulado en la hipótesis o a la magnitud del efecto. Finalmente, las variables independientes de control son las otras variables potencialmente asociadas o bien al desenlace o bien a la exposición, o a ambos, o que puedan interactuar con la variable de exposición modificando el desenlace, y cuyos efectos, por consiguiente, se quiere controlar. En la especificación de una hipótesis determinada (un modelo), suele ser necesario definir otras relaciones entre las variables de exposición, las de desenlace y las demás variables que puedan influir, como se describe a continuación.

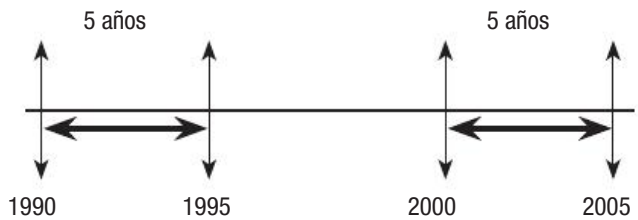
Variable de exposición

Representa la presencia y la intensidad con las que los sujetos se hallan expuestos a la variable postulada como asociada al desenlace de interés, o causante de él. Ejemplos: a) Fumar (dicótoma, Sí/No), como factor asociado al cáncer de pulmón; b) Valor promedio de HBA1c (numérica continua de razón), para predecir el riesgo de daño renal en los pacientes diabéticos.

CUADRO 4.6 Características y comparación entre niveles (escalas) de medición de intervalo y de razón

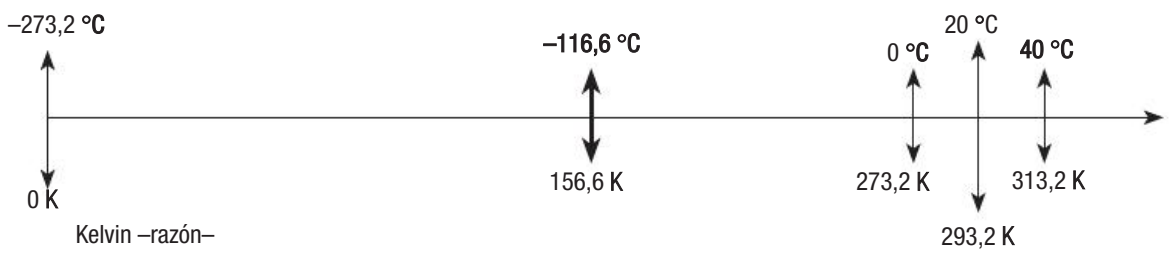
Intervalo

- La magnitud de la variable se mide con el conjunto de los números reales (positivos y negativos)
- Aplica solo a variables numéricas
  - Discretas: números enteros reales
  - Continuas: números reales
- El cero de la escala es arbitrario
  - **Ejemplo: temperatura en °C (Celsius)**
    - ✧ **0 °C = temperatura de fusión del hielo (273,2 K)**
    - ✧ **100 °C = temperatura de ebullición del agua (373,2 K)**
    - ✧ **1 °C es 1/100 de la diferencia de temperatura entre el punto de fusión del hielo y el de ebullición del agua**
    - ✧ **-20 °C no es "ausencia" o "deuda" de temperatura: equivale a 253,2 K**
- Las distancias entre valores consecutivos son simétricas
  - Las adiciones y las diferencias tienen sentido
  - **Ejemplo:** la distancia en años entre 1990 y 1995 (5) es la misma que entre 2000 y 2005 (5)



- Las razones entre valores no tienen sentido (porque el cero es arbitrario)
  - **Ejemplo**
    - 40 **°C no es *el doble* de 20** °C
    - 40 **°C = 313,2 K; la mitad es 156,6 K = –116,6** °C

Celsius –intervalo–

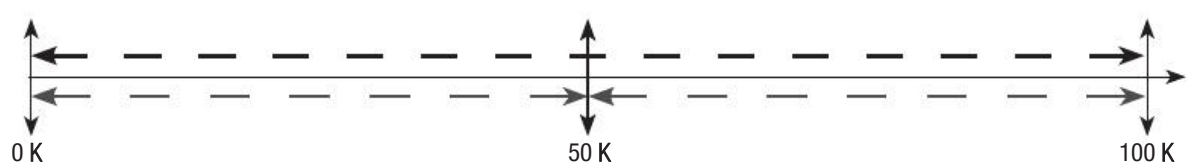


Kelvin –razón–

Razón

- La magnitud de la variable se mide con el conjunto de los números reales positivos
- Solamente aplica para variables numéricas
  - Discretas: números enteros positivos
  - Continuas: números positivos
- Al igual que el intervalo, la distancia entre valores consecutivos es simétrica
  - Sumas y restas tienen sentido
  - **Ejemplo: La distancia entre 20 K y 30 K (10) es la misma que entre 1000 K y 1010 K (10)**
- A diferencia de las de intervalo, las razones o los cocientes tienen sentido
  - **Ejemplo: 100 K es el doble de 50 K**

Kelvin –razón–



Variable de desenlace

Mide la presencia o la ausencia y la intensidad o la magnitud del desenlace de interés, del cual se postula que es el presunto resultado de la exposición al hipotético factor causal o de riesgo. Ejemplos: Cáncer de Pulmón (presente-ausente, dicótomo); Valores promedio de la hemoglobina en grupos de sujetos transfundidos con diferentes volúmenes de glóbulos rojos, por kg de peso (numérica continua, de razón).

Asociación entre exposición y desenlace

En general, la investigación cuantitativa analítica evalúa la existencia de asociaciones entre los factores (causales, de riesgo) a los que se exponen los sujetos y los desenlaces en salud. Una forma poderosa y sencilla de cuantificar dicha asociación es contrastar la frecuencia observada del desenlace de interés entre sujetos expuestos y no expuestos al putativo factor causal (variable de exposición). Una de las medidas de efecto más comúnmente empleadas es la razón o cociente entre las proporciones de sujetos con el desenlace en cada grupo de exposición, tal como se explica en detalle en el capítulo sobre medidas de frecuencia, asociación e impacto.

Si la proporción observada del desenlace en cada grupo de sujetos estudiados, definido según la presencia o la ausencia de exposición al factor de interés representa el ***riesgo*** de aparición del desenlace, el cociente entre los riesgos en pacientes expuestos y no pacientes expuestos permite cuantificar la dirección y la magnitud de la asociación entre exposición y desenlace. Este cociente se denomina ***riesgo relativo*** **(figura 4.1).**

Factor (variable) de confusión

Una variable puede desempeñar el papel de factor de confusión si cumple varios requisitos: a) está asociada independientemente al desenlace (de hecho, puede ser uno de sus factores causales); b) está asociada a la variable de exposición (se distribuye diferencialmente entre los grupos de distintos niveles de exposición), y c) no es un paso previo o intermedio en la vía causal que conecte la exposición y el desenlace.

Cuando algo de esto ocurre, la existencia del factor de confusión, efectivamente, "confunde" o distorsiona la asociación aparente entre exposición y desenlace.

EJEMPLO

Un estudio hipotético está investigando si hay asociación entre consumir café y el desarrollo de cáncer de páncreas (CAP). Al comparar las frecuencias encontradas de CAP en consumidores y no consumidores de café, se encuentra que el consumo de café se asocia a dos veces más riesgo de desarrollar CAP **(tabla 4.1).**

Al examinar si hay asociación entre el hábito de fumar y el consumo de café, los investigadores encuentran que los consumidores de café tienen cuatro veces más probabilidad de ser fumadores (*RR* = 4) **(tabla 4.2).**

Por otra parte, los fumadores muestran una frecuencia 6 veces mayor de CAP que los no fumadores **(tabla 4.3).**

Ante la incertidumbre de si el exceso de CAP en los consumidores de café es real o causado por el consumo de cigarrillo, que es más común en los tomadores de café, los investigadores estratifican (separan) su análisis, de acuerdo con la presencia o la ausencia del potencial factor de confusión, que es el hábito de fumar. Las **tablas 4.4 y 4.5** muestran las frecuencias de CAP en consumidores y no consumidores de café, en los estratos de fumadores y de no fumadores, respectivamente. Cuando se estratifica por fumar,

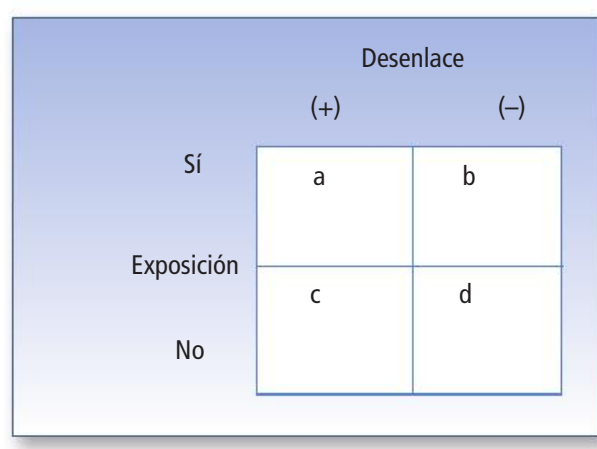


FIGURA 4.1 **Modelo de evaluación de asociación entre exposición y desenlace.**

Si la exposición se asocia al desenlace, entonces:
$A/(A + B) \neq C/(C + D)$

Tabla 4.1 Asociación "cruda" entre consumo de café y CAP

		No	
Toma café	10	990	1000
No toma café	5	995	1000
	15	1985	2000

Riesgo relativo: (10/1000) / (5/1000) = 2 veces más común el
en tomadores de café.

Tabla 4.2 Asociación entre consumo de café y el hábito de fumar

	Fuma	No fuma	
Toma café	400	600	1000
No toma café	100	900	1000
	500	1500	2000

Riesgo relativo: (400/1000) / (100/1000) = 4 veces más común el hábito de fumar en los tomadores de café.

Tabla 4.3 Asociación "cruda" entre el hábito de fumar y el CAP

		No	
Fuma	10	490	500
No fuma	5	1495	1500
	15	1985	2000

Riesgo relativo: (10/500)/(5/1500) = 6 veces más común el
en fumadores.

Tabla 4.4 Asociación entre consumo de café y CA de páncreas en el estrato de fumadores

		No	
Toma café	8	392	400
No toma café	2	98	100
	10	490	500

Riesgo relativo: (8/400) / (2/100) = 1; tan común el **en los**
tomadores de café como en los no tomadores de café.

Tabla 4.5 Asociación entre consumo de café y CAP en el estrato de los *no fumadores*

		No	
Toma café	2	598	600
No toma café	3	897	900
	5	1495	1500

Riesgo relativo: (2/600) / (3/900) = 1; tan común el **en los**
tomadores de café como en los no tomadores de café.

se aprecia que el riesgo de **CAP es idéntico entre** los tomadores de café y los no tomadores de café que fuman **(tabla 4.4), y entre los tomadores y los** no tomadores de café que no fuman **(tabla 4.5).** La aparente asociación era producto de la confusión introducida por la variable hábito de fumar, asociada independientemente a la exposición (consumo de café) y el desenlace (**CAP**).

En conclusión, el estudio hipotético analizado encontró que consumir café no aumenta el riesgo de **CAP, pero el hábito de fumar sí** puede estar asociado a dicho cáncer. Como los tomadores de café fuman más, eso da la falsa impresión de que tomar café aumenta el riesgo de **CAP. La variable fumar confundió la** asociación entre tomar café y **CAP, y su efecto** tuvo que ser controlado mediante un análisis estratificado.

Modificador de efecto

Una variable de control es un *modificador de efecto* si interactúa con la variable de exposición modificando la magnitud de su efecto sobre el desenlace.

La modificación de efecto (interacción) es diferente de la confusión, y, de hecho, ambas pueden presentarse de manera independiente o coexistir en un estudio específico. La magnitud encontrada del efecto es diferente de la adición de los efectos separados de cada una de las variables independientes (o potenciales predictoras) que interactúan.

EJEMPLO

Se halla claramente establecido que la exposición a cierto tipo de asbestos aumenta el riesgo de un tipo de cáncer pulmonar llamado mesotelioma. El hábito de fumar no tiene una asociación clara y específica a este tipo de neoplasia maligna. Sin embargo, se ha documentado que en las personas expuestas a asbestos el riesgo relativo de presentar mesotelioma es más elevado en quienes fuman que en quienes no fuman. Esto se interpreta como la existencia de una interacción entre el hábito de fumar y la exposición a asbestos, en la que, aparentemente, la exposición al humo de tabaco potencia el efecto carcinogénico de los asbestos.

Variables de control

En general, todas las variables que independientemente podrían asociarse a cambios en la frecuencia del desenlace tienen la posibilidad de distorsionar la evaluación de la asociación entre la exposición y el desenlace de interés. Durante el diseño, el investigador trata de identificar y anticiparse al posible efecto distorsionador de dichas variables. En general, estos potenciales modificadores de efecto, o factores de confusión, entran en la categoría general de ***variables de control***, que deben, como su nombre lo indica, ser controladas en algún punto del proceso de investigación. Las principales estrategias para lidiar con variables de control son: 1) la exclusión, 2) el balanceo en el ensamblaje del grupo de estudio (emparejamiento o selección estratificada) y 3) el ajuste durante el análisis. Tales estrategias son discutidas en otros capítulos del libro.

Variable intermediaria o interviniente

A veces, una variable diferente de la variable de exposición y relacionada con el desenlace no se comporta como un factor de confusión, por cuanto no es un factor con asociación independiente, sino que forma parte de la secuencia o vía causal que conecta la variable de exposición con la de desenlace. Este tipo de variable recibe el nombre de ***variable intermediaria*** o ***interviniente***, pues desempeña un papel de mediadora, por ejemplo, siendo parte del mecanismo por el cual la variable de exposición incrementa el riesgo de que se produzca el desenlace.

EJEMPLO

La diabetes ***mellitus***, tanto tipo 1 como tipo 2, tiene el potencial de causar daño glomerular. A su vez, el daño glomerular puede causar hipertensión arterial. Es aparente, entonces, que sufrir diabetes puede causar hipertensión arterial como consecuencia de las lesiones glomerulares que ocasiona (glomerulopatía diabética). En consecuencia, la glomerulopatía diabética es una variable intermediaria en la relación entre la diabetes y la hipertensión arterial.

MEDICIÓN Y FUENTES DE VARIABILIDAD (ERROR)

Cuando se busca contestar una pregunta de investigación, hay cuatro posibles explicaciones para la respuesta obtenida: 1) es incorrecta, debido a la introducción de errores sistemáticos (sesgos); 2) es incorrecta por culpa de algún error proveniente de variación aleatoria; 3) es incorrecta por combinaciones de los dos tipos de error ya mencionados, o 4) es correcta. Solo cuando se han descartado razonablemente las tres primeras explicaciones, puede tenerse confianza en la veracidad de una evidencia.

Atributos de las mediciones

Al describir y evaluar el proceso científico de medición de variables, conviene definir y comprender varios conceptos involucrados en el proceso y en el producto de la medición.

La ***veracidad*** es un atributo de una observación, un hallazgo o una inferencia, y no de un proceso de medición. Se la puede definir como el grado en el que la observación de un fenómeno concuerda con la verdadera situación de lo observado.

Para verificar dicha concordancia, se requiere un resultado obtenido por un método alterno de medición, que se pueda considerar verdadero (***patrón de oro***). Es inusual que pueda comprobarse la veracidad de una observación con un elevado nivel de certeza. En la mayoría de los casos, se puede emitir un juicio sobre la probabilidad de que un dato sea correcto. La verosimilitud ("similar a la verdad") de una medida involucra hacer un juicio acerca de la probabilidad de que una observación o una medición sean veraces. Otro término adecuado para describir el resultado de dicho juicio es *confianza*. Qué tanta confianza se tiene en la veracidad de una observación, en ausencia de una verificación, contra un patrón de oro depende, en buena parte, de qué tanta confianza inspiren los métodos utilizados para hacer la medición o la observación.

Mientras que una observación o una medición pueden ser correctas (suficientemente aproximadas al valor real) o incorrectas, el

método de medición tiene una serie de atributos relacionados con la confianza que se tiene en la veracidad de las mediciones que dicho método produce.

En primer lugar está la *exactitud* (*accuracy,* en inglés): grado en el que un proceso de medición es *válido* y *confiable*, y en el que su resultado tiene una buena probabilidad de reflejar el verdadero valor de una variable.

Es entonces necesario definir de forma precisa qué se entiende por validez y por confiabilidad de una medición. El término ***validez*** se refiere a que el proceso de medición esté libre de sesgo. Un método o un instrumento de medición válidos son, respectivamente, un método o un instrumento bien calibrados. Un método está bien calibrado si al controlarse la fluctuación aleatoria de la medición (por ejemplo, repitiendo muchas veces la medición de la misma variable en el mismo sujeto), el valor promedio de las mediciones es idéntico o muy cercano al verdadero valor de la variable.

El otro atributo importante del proceso de medición es su *confiabilidad*. Un método es confiable si los resultados de las mediciones son reproducibles. Otra forma de expresar este concepto es con la palabra *consistencia*.

Un método confiable, repetible, ***o consistente***, produce observaciones precisas. La precisión se estima como el rango de variación o de fluctuación de mediciones repetidas de la misma variable, hechas con el mismo instrumento. Mientras menor sea el rango de variación, el instrumento de medición es más preciso; o sea, más ***confiable***.

Recapitulando, la validez y la confiabilidad son atributos del proceso o el método de medición, y no del resultado de la observación. Se dice que un proceso o un método de medición son ***exactos*** en la medida en la que sean simultáneamente válidos (bien calibrados) y confiables (precisos o reproducibles). El resultado de la medición (la observación en sí) no se puede clasificar como válida y confiable, sino como acertada (veraz) o incorrecta.

Se espera, entonces, que los procesos de medición capturen la verdadera variabilidad en las características observadas o medidas (variables medidas), y que no se deje distorsionar por una variabilidad que no corresponde al fenómeno medido, sino a otros factores. La variabilidad extraña, que no es inherente al verdadero estado de la variable medida, se conoce como ***error***. El error, en tal sentido, no necesariamente quiere significar que la medición es equivocada, sino señala el efecto de las limitaciones del proceso de medición en cuanto a su validez y su confiabilidad.

Fuentes de error en la medición

Las fuentes de variación en la medición de una variable son tres: 1) el sujeto o el fenómeno medido, 2) el instrumento utilizado para medir y 3) el observador que hace la medición.

Los cambios provenientes del sujeto medido pueden reflejar la variabilidad de los elementos de un conjunto; es decir, la variación que exhiben en una población los sujetos respecto a una o varias características (variación interindividual), o bien, originarse en los cambios ocurridos en un mismo sujeto a lo largo del tiempo (variación intraindividual).

EJEMPLO

La frecuencia cardíaca de cada uno de los individuos en un grupo de pacientes puede (y suele) ser diferente (variación interindividual). Además, la frecuencia cardíaca medida en un mismo paciente en dos periodos diferentes también suele ser distinta (variación intraindividual). Esta variación hace que a quien efectúa la medición le quede difícil determinar cuál es la frecuencia cardíaca "verdadera" o "típica" de un grupo, y, posiblemente, deba consignar un rango de valores observados a menudo, para acoger la variación tanto interindividual como intraindividual.

Los instrumentos tienen limitaciones para medir los fenómenos y discriminarlos (la "resolución", o capacidad discriminativa); además, pueden distorsionar la verdad de los hechos (cometer errores) de dos maneras: 1) sistemática y 2) aleatoria.

Finalmente, el observador tiene sus propias limitaciones y prejuicios cuando utiliza el instrumento de medición o cuando interpreta las observaciones.

Tipos de error en la medición

Independientemente de cuál sea la fuente de variabilidad en la medición, el error se puede presentar de dos maneras: 1) desviación sistemática y 2) fluctuación aleatoria.

✔

La ***desviación sistemática*** de las mediciones (en una misma dirección) es un tipo de error que afecta la validez de las mediciones, y se conoce como ***sesgo***; la ***variación*** aleatoria, o azar, afecta la precisión de las mediciones. Sin embargo, los dos tipos de error pueden coexistir en una medición.

Sesgo

Por ***sesgo*** entendemos una desviación sistemática de la realidad de los hechos o de los fenómenos; debido a ello, los sesgos pueden operar en cualquier punto de la elaboración de una investigación.

Un investigador quiere evaluar la relación entre el hábito de fumar y la presencia de carcinoma de vejiga. Para tal propósito, hace un estudio de cohorte donde compara las causas de mortalidad de dos grupos de pacientes, fumadores y no fumadores, observados durante un periodo prolongado (30 años). La causa de los fallecimientos se detecta por medio de los certificados de defunción; sin embargo, en su afán por mejorar la detección del cáncer de vejiga, el investigador decide revisar las historias clínicas de pacientes y complementar los informes de los certificados de defunción con revisiones de autopsias y entrevistas a los médicos tratantes. Como dispone de recursos limitados y juzga que la información relativa a los fumadores es más importante, limita al grupo de fumadores estas maniobras de búsqueda activa del cáncer. Debido a que las maniobras para demostrar cáncer de vejiga son más exhaustivas en los pacientes expuestos al cigarrillo que en los no expuestos, la probabilidad de detectar el desenlace de interés es sistemáticamente mayor en los fumadores, y su medición del riesgo de cáncer asociado al hábito de fumar está sesgada. En caso de hallarse asociación entre tabaco y carcinoma de vejiga, será muy difícil saber si la asociación es real o si ha sido provocada por la diferencia sistemática (sesgo) entre los procedimientos de medición de frecuencia empleados en los dos grupos estudiados.

✔

Existen tres grandes tipos de sesgos: 1) de selección, 2) de confusión y 3) de medición.

Sackett ha catalogado más de 35 tipos posibles de sesgos provenientes del muestreo y de la medición, aunque todos ellos pueden estar dentro de las tres grandes categorías mencionadas. Las otras circunstancias en las cuales operan los sesgos (la selección de sujetos, la interpretación de datos, etc.) donde no está directamente involucrado un error en la medición de una variable son discutidas en otros capítulos.

Los estudios de investigación que involucran a seres humanos son más propensos a ser afectados por sesgos. La evaluación del papel que desempeña un sesgo es importante en el análisis de los datos —aunque su cuantificación sea difícil—, porque es fundamental, dentro del diseño de un estudio, anticiparse a la presencia de un sesgo para introducir estrategias que lo eviten.

✔

En el proceso de medición, el sesgo se produce cuando el proceso (el sujeto medido, el instrumento de medición o el observador) está "descalibrado" y los resultados de las mediciones se desvían consistentemente en una misma dirección, y se alejan así del valor real.

Variación aleatoria

Siempre que se registra información, para la práctica clínica o en el curso de una investigación, se está, de alguna manera, midiendo una variable en una muestra de una población (teórica o real) de posibles valores de dicha variable. Si se considera el resultado de un hemograma, por ejemplo, uno se enfrenta a los valores obtenidos en una muestra de unos 5 ml, con los cuales se quiere estimar esos mismos valores ***en la totalidad de la sangre circulante*** en una persona. La presunción es que la muestra obtenida es, de cierta forma, "representativa" (típica) del resto de la sangre que no pudimos examinar; o sea, que al recolectarla no se incurrió en ningún sesgo (desviación sistemática).

Cuando el clínico interpreta los resultados del examen, siempre considera la posibilidad de la presencia de sesgo en la medición. Por

ejemplo, si la sangre se obtiene mediante una línea intravenosa y está diluida con la solución que se le esté administrando al paciente. Pero si se descarta dicho tipo de origen de variación, siempre cabe la posibilidad de que el valor no sea representativo, pues se tuvo la "mala suerte" de tomar una muestra atípica. Probablemente, eso sucede pocas veces, dado que la sangre es un coloide con concentraciones muy uniformes de partículas.

Al aceptar la anterior premisa, se está aceptando que la distribución teórica de los valores de todas las muestras posibles de sangre por examinar en un cuadro hemático es definible, y que se puede calcular la probabilidad de que al tomar una muestra no sesgada de sangre dicha muestra sea una adecuada representación de los valores sanguíneos en el individuo examinado. La teoría de probabilidades permite estimar el posible papel del azar en el resultado de una medición específica.



En general, debido a la incertidumbre aleatoria inherente a todo proceso de medición, cuando se obtiene un valor específico en una medición, dicho valor es compatible con un intervalo de valores cuyo rango es proporcional a la reproducibilidad del proceso. El rango de valores consistentes con la medición efectuada es lo que se conoce como un *intervalo de confianza*. Cuanto más preciso sea un proceso de medición, tanto más estrecho será el rango de dicho intervalo.

Referencias

1. Chobanian AV, Bakris GL, Black HR, et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension. 2003;42:1206-52.

Lecturas recomendadas

Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: The essentials. 5th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2014.

Kirkwood B, Sterne J. Essential medical statistics. 2nd ed. Malden, MA: Blackwell Science; 2003.

Kleinbaum D, Kupper L, Morgenstern H. Epidemiologic research. Principles and Quantitative methods. 1st ed. Nueva York: Van Nostrand Reinhold; 1982.

Norman GR, Streiner DL. Biostatistics: The bare essentials. 3rd ed. Philadelpia: Peoples Medical Publishing House; 2008.

Penrose R. El camino a la realidad: Una guía completa a las leyes del universo. 1ª ed. Madrid: Editorial Debate; 2006.

Straus SE, Glasziou P, Richardson WS, et al. Evidence based medicine, 4th Ed. Edinburgh: Churchill Livingstone; 2010.

CAPÍTULO 5

Población y muestra

Juan Gabriel Ruiz Peláez • María Nelcy Rodríguez Malagón

INTRODUCCIÓN

Generalidades

Uno de los objetivos fundamentales de la investigación científica en salud es aportar conocimientos válidos y precisos como insumo fundamental para la toma de decisiones racionales en la atención en salud, de cualquier clase: desde el cuidado específico de pacientes y sus problemas hasta la formulación de estrategias y de políticas sociales de prestación de servicios en salud. En particular para el clínico que necesita basar sus decisiones asistenciales en conocimiento científico, es crucial que los resultados de la investigación sean válidos, confiables y aplicables.

Para poder capturar confiablemente y de manera no sesgada la variabilidad en las respuestas y las asociaciones entre variables al estudiar salud y enfermedad, así como sus condicionantes y la racionalidad y la utilidad de potenciales intervenciones, debería estudiarse a la totalidad de los sujetos (***elementos***) que constituyen la población o el conjunto de interés.

En la vida real, no obstante lo anterior, las observaciones y las mediciones que se requieren para obtener información suficiente, que permita llegar a conclusiones válidas, rara vez pueden hacerse en el total de la población de interés. Hay múltiples razones que limitan la factibilidad para hacer mediciones en todos los sujetos: limitaciones en el acceso a recursos económicos y de tiempo para lograr un cubrimiento de la totalidad de una población definida; la escasez de recursos humanos calificados para hacer las observaciones o para obtener las mediciones necesarias, y las dificultades geográficas y logísticas que presenta una tarea de gran magnitud como el cubrimiento total de una población geográficamente definida. Adicionalmente, la población de interés puede ser ilimitada (por ejemplo, el conjunto de las personas con enfermedad coronaria), en cuyo caso, el cubrimiento completo de la población de interés es un imposible lógico.

Frente a estos problemas conceptuales y prácticos, se ha planteado el ***estudio de muestras*** (conjuntos finitos y definidos de elementos) tomadas de la población de interés. Si este subconjunto representa adecuadamente la composición y la variabilidad de los elementos integrantes de la ***población de interés*** (el conjunto de donde se extrae la muestra), las observaciones y las conclusiones obtenidas pueden utilizarse para generalizar los resultados a la población de interés (cuando la muestra es probabilística de una población finita y definida en tiempo y espacio; ver más adelante la definición de *población blanco*) o para extrapolar (cuando la muestra es un modelo de la composición y relaciones en una ***población abstracta***, o ***de referencia***, también discutida más adelante).

El proceso mediante el cual las observaciones y los análisis realizados se conectan en la muestra a la estimación de parámetros o a la evaluación de hipótesis relevantes para la población de interés se denomina ***inferencia***.

Preguntas de investigación, poblaciones y muestras

En el **capítulo 2** ("La pregunta de investigación") se ha reflexionado sobre los elementos de una pregunta de investigación apropiada, que se constituyen en un paso fundamental, necesario, aunque no suficiente, para la adquisición de conocimiento de forma sistemática, válida y relevante. Este capítulo se centrará en los elementos o los sujetos a quienes es necesario estudiar (observar), para tratar de contestar una pregunta de investigación científica.

Hay dos matices importantes en la definición de la validez de un proceso de investigación cuantitativa, clínica o epidemiológica: a) ¿son las observaciones y los procesos de investigación adecuados y veraces en los sujetos estudiados? (***validez interna***), y b) dado que los resultados o las conclusiones son válidos ¿a quiénes pueden aplicarse? (***validez externa***).

Este capítulo se enfoca, principalmente, en la descripción y el análisis de los componentes, los procesos y los problemas relacionados con la validez externa de una investigación. Se revisarán los conceptos relacionados con poblaciones y muestras: selección, muestreo y asignación de sujetos, sesgo y azar.

En el enfoque cuantitativo hay dos tipos fundamentales de preguntas: las ***descriptivas*** y las ***analíticas*** (véase el **capítulo 2**, "La pregunta de investigación", y el **capítulo 13**, "Introducción a los diseños de investigación: selección del diseño apropiado"). Los conceptos y las consideraciones relacionados con los sujetos a quienes se va a estudiar varían y tienen aspectos específicos según el tipo de pregunta.

DEFINICIONES Y CONCEPTOS

Población y muestra en preguntas descriptivas

En el caso de las preguntas descriptivas, se quiere conocer la distribución de una o de varias características de un conjunto de elementos, conocido como ***población***. Una población no necesariamente está constituida por seres humanos o seres vivos, sino que corresponde al concepto matemático de ***conjunto***: un grupo de elementos definido de forma precisa. ¿Cómo averiguamos la distribución de una o más variables en un conjunto específico? Hay al respecto dos alternativas: 1) examinamos todo el conjunto; es decir, hacemos un censo donde se busca evaluar a la totalidad de los elementos del conjunto en un momento dado, o examinamos solo una porción de los elementos (o sea, una *muestra*), y con base en las mediciones obtenidas en la muestra intentamos establecer el valor de dichas variables en todo el conjunto.

Población blanco

Al conjunto al que se le quiere medir la distribución de variables se lo conoce como ***población blanco***, ***objetivo*** o ***diana***. El valor de la variable que se quiere medir en la población blanco se denomina ***parámetro***. Por ejemplo, la frecuencia de diabetes ***mellitus*** tipo II en todas las personas afiliadas a una entidad prestadora de salud (EPS) específica, en una ciudad también específica (Buriticá) y en una fecha determinada (mayo de 2010), es el parámetro que se quiere estimar. La variable de interés puede ser, por ejemplo, numérica continua (glucemia en ayunas), y en ese caso se puede querer estimar la distribución de la variable o estimar algunos índices que resuman dicha distribución: el promedio de las glucemias, la desviación estándar, etc. En tal caso, la medida de resumen de la distribución medida en toda la población blanco, (por ejemplo, el promedio) es el *parámetro poblacional*. Por supuesto, la variable también puede ser discreta, como, por ejemplo, la proporción de sujetos con hiperglucemia en ayunas, o la prevalencia de la hiperglucemia.

Muestra

Cuando se selecciona un grupo de elementos (muestra) y en ellos se mide la variable de interés, lo que se obtiene respecto a la población blanco es un estimador del parámetro poblacional. Por ejemplo, el valor de la media de glucemia en ayunas en una muestra de 50 afiliados a EPS en Buriticá en mayo de 2010 es el ***estadígrafo***, o estadística que se obtiene, y se usa para estimar el parámetro (el promedio de glucemia en todos los afiliados residentes en Buriticá en mayo de 2010) en la población blanco.

En resumen, el problema general que se debe resolver para contestar preguntas descriptivas consiste en estimar parámetros de poblaciones concretas (poblaciones blanco) a través del examen de muestras. Para que las muestras permitan hacer una estimación creíble y confiable (o sea, válida y precisa), la muestra tiene que ser representativa de la población blanco.

Validez

Una muestra es válida si permite estimar de forma no sesgada los parámetros de interés. La única manera de asegurar la validez del proceso de muestreo es que los elementos incluidos en la muestra sean escogidos mediante un proceso aleatorio. De esta forma, cada uno de los elementos del conjunto tiene la misma oportunidad de ser incluido, y se espera que si hay una característica (por ejemplo, ojos cafés) que comparta el 80 % de los elementos de la población blanco, alrededor del 80 % de los elementos incluidos en la muestra escogidos al azar tenga ojos cafés. El muestreo aleatorio equivale a homogeneizar la concentración, tal como ocurre con una sustancia soluble que se encuentra en una solución: la concentración de moléculas de la sustancia en cualquier volumen dado de la solución es idéntica, de suerte que si uno examina una muestra de cualquier parte de la solución debe obtener la misma concentración.

Precisión

La precisión de la estimación hecha con una muestra aumenta a medida que se incrementa el número de elementos incluidos. Si se incluyen todos los elementos (***censo***), la muestra se hace igual a la población blanco, y la estimación del parámetro es totalmente válida y precisa. Sin embargo, en muchos casos medir en toda la población es impráctico o imposible: por ejemplo, para estimar la proporción de fósforos que enciende correctamente en una caja de ellos, no encendemos todos los elementos, sino una muestra.

Tamaño de muestra para la estimación de parámetros de una población blanco

Una cuestión muy importante es decidir cuántos elementos del conjunto hay que incluir; o sea, cuál debe ser el tamaño de la muestra. En esta sección del presente capítulo no se enumeran ni se discuten las fórmulas empleadas para calcular el tamaño muestral, pero sí se mencionan los factores que influyen en esa decisión: la variabilidad de la característica de interés en la población blanco y la precisión que se quiere obtener en la estimación (el llamado *error muestral*).

Supóngase que se quiere calcular la estatura promedio de los alumnos de Décimo Grado de un colegio en particular, en una fecha determinada (población blanco). En el caso extremo de que ***todos*** los alumnos tengan exactamente la misma estatura, bastaría con examinar a un sujeto para conocer la estatura de todos los demás. Esto ocurre porque la variabilidad de la característica en la población blanco es cero. Si, por el contrario, las estaturas de los sujetos varían mucho, se necesita incluir a muchos sujetos. Por ejemplo, asumamos que el 20 % de los alumnos mide 166 cm; el 20 %, 167 cm; el 20 %, 170 cm; el 20 %, 175 cm, y el último 20 %, 178 cm; es decir, que hay 5 subconjuntos de talla. En este caso no se puede incluir a menos de 5 sujetos en la muestra, pues si hay, por ejemplo, solo 4, es imposible que los 5 grupos de estaturas queden representados.

Ahora bien, con solo cinco sujetos escogidos al azar no es improbable que únicamente salgan representados dos o tres de los cinco grupos de estatura. La confianza que se tendría en dicha estimación sería muy baja (***error muestral grande***). Si se incluye a más sujetos, aumentan las probabilidades de que la distribución de las tallas se asemeje más a la de la población blanco (el error muestral disminuye). Antes de hacer una investigación descriptiva, el investigador debe determinar cuánto error muestral está dispuesto a tolerar, y debe tener alguna idea fidedigna de la variabilidad de la característica en la población objetivo, y con esa información, emplea una fórmula matemática, la cual le arroja como resultado el número de elementos que es necesario incluir en la muestra. En síntesis: a mayor precisión deseada, mayor tamaño de muestra, y a menor variabilidad en la población blanco, menor tamaño de muestra.

Población blanco, población de estudio, población accesible y muestra

Recapitulando, hay muchas instancias en las que no es posible o no es práctico examinar todos los elementos de una población, y debe-

mos limitarnos a examinar una muestra. Por ejemplo, para estimar el contenido de hemoglobina en la sangre de una persona, tomamos una muestra representativa, ***no la totalidad*** *de la sangre*. Debido a que la sangre es un coloide con partículas y células suspendidas de forma homogénea, al tomar una muestra de una vena de tamaño medio, la concentración de hemoglobina o de eritrocitos es uniforme y la medición de dicha concentración debe reflejar muy estrechamente la del resto de la sangre que circula por esa vena. De hecho, en fisiología del sistema circulatorio se enseña que la concentración de las células sanguíneas en todos los vasos venosos de mediano y gran calibre es muy semejante.

En el problema de estimar la concentración de hemoglobina, la población blanco se halla constituida por toda la sangre del cuerpo del sujeto, en un momento dado. En realidad, no se tiene un acceso expedito y seguro a muestras de cualquier compartimiento sanguíneo. La sangre circula por diferentes vasos (arterias, venas) de diferente calibre (grandes, medianos, pequeños, vasos capilares) y a diferentes velocidades, y su concentración puede variar: en el lecho capilar, la sangre está ligeramente más concentrada que en el lecho arterial o en el venoso. Generalmente, obtenemos las muestras de sangre al puncionar una vena de mediano calibre en una extremidad. La sangre que circula por dichas venas es relativamente homogénea y ligeramente distinta de la capilar. La sangre de dichas venas puede ser la población elegible (para muchos autores, también llamada población de estudio); o sea, el subconjunto de la población blanco que, efectivamente, se va a evaluar. La población elegible, o de estudio, viene definida por la población blanco (el tiempo es ahora, el lugar es la sangre venosa y la persona es la concentración de hemoglobina del sujeto determinado); y también está definida por criterios de inclusión, pues se examina sangre anticoagulada y de exclusión, porque no sirve si la vena está trombosada. Ahora bien, no todas las venas son igualmente fáciles de puncionar. De la sangre que viaja por las venas periféricas (es decir, de la población elegible, o de estudio) solo es potencialmente accesible la que circula por determinadas venas. Esa constituye el subconjunto que se conoce como ***población accesible***: esa fracción poblacional elegible, y a la cual, efectivamente, se tiene acceso para incluir en la muestra.

En síntesis, cuando se observa con cuidado un problema de estimación en apariencia simple, como calcular la concentración de hemoglobina en la sangre venosa en un paciente, se hace notorio que la inferencia de la muestra hacia la población objetivo no es directa. La muestra se toma de una población fuente, y si el muestreo es no sesgado y preciso, el estimador puede representar de forma válida y confiable al parámetro en dicha población fuente. La situación ideal es cuando la población fuente es directamente la población blanco; o, dicho de otro modo, cuando la población blanco, la elegible y la accesible son la misma. En tal caso, se generaliza un resultado obtenido en una muestra a toda la población.

Cuando lo anterior no ocurre (como en el ejemplo de la muestra de sangre), el resultado solo se puede generalizar a la población accesible (el resto de sangre en la vena puncionada), y si no hay razones para pensar que dicha población accesible sea sistemáticamente diferente de la población de estudio, o elegible (sangre venosa de venas periféricas), los resultados se pueden extrapolar a dicha población elegible. A su vez, si la población elegible no es una representación distorsionada de la población blanco (toda la sangre del sujeto en ese momento), se extrapola, finalmente, a dicha población objetivo. En este ejemplo, es claro que los resultados no son extrapolables a toda la sangre, pues la que está en los lechos capilares (y forma parte de la población blanco, pero no de la población elegible) tiene concentraciones de hemoglobina ligeramente más elevadas.



Para recordar, el resultado obtenido en la muestra se generaliza solo a la población fuente (población accesible). La inferencia a las otras poblaciones de la cadena inferencial es ***extrapolación***.

Población y muestra en preguntas analíticas

El problema al que nos enfrentamos puede ser descrito así:

Se ha planteado una hipótesis, que no es sino una conjetura acerca de cómo funciona la realidad. A partir de esa conjetura, se han deducido las consecuencias (***predicciones***) que tendría dicha hipótesis, en caso de ser cierta.

Por ejemplo, se plantea la hipótesis de que el tratamiento oportuno de las infecciones de piel (impétigo) causadas por cepas nefritógenas de *Streptococcus pyogenes* con antibióticos eficaces disminuye el riesgo de aparición de glomerulonefritis postestreptocóccica (GPE). Al hacer el ejercicio de identificar los elementos de la pregunta analítica, se encuentra lo siguiente:

1. **Dominio:** Prevención (mediante el tratamiento de una condición, prevenir la aparición de otra).
2. **Población:** Sujetos previamente sanos (especialmente, sin enfermedad renal), con impétigo causado por una cepa nefritógena de estreptococo β hemolítico grupo A (EBHGA).
3. **Exposición (intervención):** Administración de penicilina (por ejemplo, penicilina V oral, a dosis adecuadas, por 10 días), iniciada dentro de las primeras 48 horas de aparición del impétigo.
4. **Control (intervención de control):** Manejo local del impétigo, con descostrado y antibacteriano tópico (ácido fusídico).
5. **Desenlace:** Ocurrencia de GPE (incidencia).

Para contestar la pregunta planteada sobre el tratamiento para las infecciones de piel, se deben hacer observaciones de la realidad (*recoger datos*), para ver si las predicciones hechas por la hipótesis se cumplen. En principio (ateniéndonos a la propuesta de Popper), se quiere recolectar información que permita rechazar la hipótesis (la *falsabilidad*, como arma de la investigación científica). En este ejemplo en particular, se manipula a los sujetos (unos reciben penicilina, y otros, no) y se observan los resultados.

¿Población o poblaciones?

En realidad, hay dos subpoblaciones "blanco": la de sujetos expuestos a la penicilina y la de no expuestos a la penicilina; y la pregunta clave radica en saber si el parámetro de interés (frecuencia de GPE) es el mismo para ambas subpoblaciones (es decir, si son, en términos prácticos, la misma población) o si, por el contrario, se trata de poblaciones diferentes con riesgos distintos. La respuesta se obtiene al estimar la frecuencia de GPE en ambos grupos. Se trata de establecer si los sujetos con impétigo, por el hecho de recibir penicilina, ya no pertenecen a una población con un riesgo elevado de desarrollar GPE, sino que, gracias a la acción del antibiótico, ahora pertenecen a una población diferente, con un riesgo más bajo.

No podemos hacer estas observaciones en "toda la realidad". Es imposible recabar información de todos los sujetos para quienes sea pertinente la hipótesis planteada o la pregunta de investigación formulada. Por tanto, tenemos que desarrollar una aproximación sistemática y válida al problema de seleccionar solo a algunos sujetos en quienes hacer nuestras observaciones.

Población de referencia *vs.* población blanco

Si se examina con cuidado la población "blanco" del ejemplo, se nota que difiere de forma significativa del concepto de población blanco, empleado al discutir las preguntas descriptivas. La población blanco en una pregunta descriptiva puede ser grande o pequeña, pero está claramente definida en cuanto a tiempo, lugar geográfico y personas: sujetos afiliados a EPS en Buriticá en mayo de 2010. Por el contrario, la población de interés en el ejemplo del impétigo está definida de forma *abstracta* (sujetos con impétigo por cepa nefritógena, sin enfermedad renal), sin precisar tiempo (podría ser durante la prehistoria o en el futuro lejano) ni lugar (cualquier comunidad donde vivan seres humanos y donde haya EBHGA nefritógeno). Solo las características de las personas (demografía, sujetos previamente sanos, con impétigo por EBHGA nefritógeno, etc.) están rigurosamente definidas.

Población de referencia

A diferencia de la población blanco en Buriticá, que tiene un número finito y mensurable de elementos, la presente población es o bien potencialmente infinita, o, al menos, ilimitada. En realidad, se trata de una abstracción, de un concepto que se les puede aplicar, a manera de definición, a distintos sujetos pertenecientes a muchas poblaciones blanco específicas. A esta población abstracta, definida solo en términos de "persona" (características de los elementos que satisfacen una definición) se la conoce como *población de referencia*. Los ejemplos abundan: población de diabéticos, de diabéticos hipertensos, de pacientes con neumonía (los sujetos entran en la población de referencia cuando enferman, y salen cuando se curan), etc. Dado

que se trata de poblaciones abstractas, con un número potencial infinito o ilimitado de elementos, es matemáticamente imposible determinar sus parámetros: los denominadores de cualquier proporción o de cualquier sumatoria para promedios serán infinitos, y las fracciones calculadas serán siempre cero. Este concepto matemático se corresponde con el concepto epistemológico de imposibilidad de hacer afirmaciones generales con base en observaciones particulares, conocido como la ***falacia de la inducción***, o *afirmación del consecuente*.

Población de prueba e inferencia (generalización *vs.* extrapolación)

A pesar de la mencionada imposibilidad matemática de determinar parámetros en poblaciones de referencia, cuando generamos conocimiento científico (descripciones, predicciones) estimamos parámetros y establecemos las relaciones matemáticas que, aparentemente, "gobiernan" (leyes, con fórmulas derivadas de observaciones e hipótesis) o "explican" (teorías que incluyen e interpretan dichas "leyes") los fenómenos que ocurren en esas poblaciones abstractas. En realidad, generamos modelos en muestras de sujetos que satisfacen la definición de la población de referencia, para tratar de hacer inferencias (predicciones o explicaciones) que se apliquen a otros sujetos que también pertenezcan a la población de referencia. Se trata de *inferencia inductiva*, **un esfuerzo de inferir** de lo particular a lo general.

Así la pregunta haya surgido de una población concreta (mis pacientes con impétigo), se busca que la respuesta no simplemente sea aplicable a unos pacientes determinados, sino a cualquier miembro de la población de referencia (cualquier paciente con impétigo, previamente sano, etc.). La muestra no necesariamente tiene que ser representativa de mis pacientes, siempre y cuando satisfaga adecuadamente la definición que le permite pertenecer a la población de referencia.

De hecho, como lo que se construye es un modelo para observar o para experimentar, los sujetos se incluyen pensando en que permitan contrastar de forma adecuada las observaciones, no que necesariamente sean "típicos" representantes de una población blanco específica. Este concepto es llamado población de prueba por Greenland. La población fuente de los sujetos siempre será, de forma abstracta, la población de referencia que haya sido definida. De forma concreta, los pacientes pueden venir de la misma población blanco donde se ha documentado el problema (el riesgo de impétigo) y a la que se quiere intervenir, o bien, provenir de sujetos totalmente diferentes, pero que pertenecen a la población de referencia (la ya mencionada población de prueba). El concepto se puede extender incluso a poblaciones de animales que representen adecuadamente el problema humano que se quiere estudiar. Si ese concepto no fuera extensible, la experimentación básica en modelos de tejidos y en animales no sería utilizable para aprender y tratar enfermedades humanas.

No obstante lo anterior, desafortunadamente, el proceso de inferencia no siempre es tan directo y claro. En realidad, el investigador trabaja con una población concreta de sujetos que satisfacen la definición de la población de referencia (población de estudio, o elegible), pero que puede tener peculiaridades (criterios de inclusión y de exclusión) adecuados para que la investigación sea sólida, pero que pueden dificultar la extrapolación a sujetos que no pertenecen a la población de estudio, pero sí tienen el problema de interés (y, por tanto, pertenecen a la población de referencia). Más aún, tal como se discutió en el aparte sobre población en estudios descriptivos, de todos los sujetos potencialmente elegibles, el investigador puede tener acceso tan solo a unos pacientes específicos, y no necesariamente representativos de todos los sujetos de la población de estudio.

En el proceso de inferencia se evalúa la hipótesis en los sujetos que componen la muestra, y si dicha evaluación es válida, se tiene confianza en la veracidad del resultado para los sujetos estudiados (***validez interna***). Se podría seguir la cadena de inferencia de forma similar a como se hace con investigación descriptiva: si los sujetos incluidos son una representación no sesgada de la población elegible, o de la población de estudio, uno podría generalizar a dichas poblaciones fuente.

Pero recordemos que, en un sentido abstracto, la verdadera población fuente es la población de referencia, en la cual, matemáticamente, es imposible determinar ningún parámetro (porque los denominadores son infinitos o ilimitados). El problema aquí, tal como ya se mencionó, no es de generalización, sino, simplemente, de

extrapolación a otros sujetos que satisfagan la definición de población de referencia. La muestra estudiada no busca ser representativa de una población blanco, sino ser un modelo adecuado para describir las relaciones (predecir o explicar) de variables en la población de referencia.

POBLACIÓN ELEGIBLE (DE ESTUDIO); CRITERIOS DE INCLUSIÓN Y DE EXCLUSIÓN

En principio, la muestra que se va a estudiar es, simplemente, un grupo de elementos que pertenecen al conjunto "población de referencia"; por ejemplo, sujetos con impétigo por EBHGA nefritógeno. Sin embargo, por ser la población de referencia una población abstracta, de ella no se toma directamente la muestra. En el mundo real, la muestra proviene de una población fuente concreta, definida en tiempo, lugar y características de los sujetos (personas): por ejemplo, pacientes de la consulta de dermatología de un hospital universitario (lugar) diagnosticados y seguidos a partir de mayo de 2010 (tiempo) y que tengan diagnóstico de impétigo por EBHGA (persona). A menudo, la definición de las características personales de los sujetos debe ser mucho más clara y específica: que el diagnóstico de infección por EBHGA sea hecho con criterios estrictos, válidos y reproducibles, etc.

Adicionalmente, los estudios analíticos buscan establecer si existe asociación entre la exposición a un factor (por ejemplo, tratar el impétigo con penicilina V oral) y un desenlace determinado (por ejemplo, daño renal por GPE); y lo logra al comparar la frecuencia del desenlace cuando se está expuesto al factor con la frecuencia cuando se tiene una exposición de control; por ejemplo, una exposición diferente o habitual: usar solo descostrado y limpieza de las lesiones de impétigo.

En el caso descrito, la población elegible está compuesta por sujetos que pertenecen a la población de referencia (tienen impétigo), son accesibles al investigador (población accesible en tiempo y lugar determinado) y, además, satisfacen unos requisitos especiales que permiten evaluar la hipótesis de asociación. Los dos requisitos indispensables son: a) que al sujeto le pueda ocurrir el desenlace de interés; es decir, que sea susceptible (tiene impétigo y podría sufrir daño renal), y b) que en el momento de ser elegido para el estudio no tenga el desenlace de interés (que no tenga GPE y tenga su función renal normal).

Las mencionadas condiciones permiten que se pueda estudiar la hipótesis de asociación entre el factor de interés (antibiótico *vs.* manejo habitual con descostrado) y el desenlace (protección contra la GPE —daño renal—), como se aclarará más adelante.

Selección y muestreo

Si un sujeto de la población accesible no cumple al menos con los dos requisitos del caso, no pertenece a la población elegible y no debe ser incluido en la muestra de estudio. Se suelen especificar otros requisitos que caracterizan de forma muy definida a la población elegible, y se enumeran como *criterios de inclusión y de exclusión*. Los criterios de inclusión explican características que debe tener el sujeto para ser incluido (por ejemplo, edad entre 4 y 18 años) y los de exclusión enumeran motivos por los cuales, a pesar de ser elegible, no debe ser incluido (por ejemplo, ser alérgico a la penicilina). Los sujetos que cumplen todos los requisitos de admisión son potencialmente elegibles. Estos sujetos así definidos se asemejan al concepto de población de referencia, en el sentido de que solo se hallan caracterizados en términos de quiénes son, qué padecen, de cuáles complicaciones están libres y qué tan adecuados resultan para el estudio, pero no han sido definidos en términos de tiempo y lugar. Como el estudio debe llevarse a cabo en una población concreta, es necesario definir un grupo de sujetos elegibles (*abstracto*), que, además, sean accesibles (*concreto*). Determinar así la población elegible que, además, es accesible, se denomina ***selección***.

De entre esos sujetos se escoge una ***muestra***, un conjunto de individuos provenientes de la población elegible y accesible (población concreta en el mundo real), que, a su vez, pertenecen a una población de referencia (población abstracta). Es esa relación la que se invoca cuando se quiere extrapolar lo aprendido en los estudios a otros sujetos que satisfacen la definición de la población de referencia, pero no pertenecen a la muestra estudiada ni a su población fuente. Extraer a un grupo de sujetos de la población seleccionada como elegible es la toma de la muestra, o ***muestreo***.

El grupo de sujetos que se ha seleccionado e incluido en el estudio es lo que muchos epidemiólogos reconocerían como una ***cohorte***: un conjunto homogéneo de sujetos, que en el momento de ensamblaje de la cohorte es susceptible de presentar un desenlace (daño renal por GPE por padecer impétigo), pero que aún no lo ha presentado, y puede o no exponerse al factor de interés (recibir o no penicilina V). El seguimiento de una cohorte permite estimar la incidencia del evento de interés, y es un grupo ideal para contestar preguntas analíticas, bien sea con estrategias ***observacionales*** (estudios de cohorte) o ***experimentales***. Cuando se selecciona la muestra, se está ensamblando la cohorte que va a ser estudiada. Incluso en el caso de los estudios de casos y controles, donde se parte de la presencia (caso de GPE) o de la ausencia (control, sin GPE) del desenlace, se asume que tanto casos como controles provienen de una cohorte subyacente (tomada de una población fuente conformada por sujetos quienes en el pasado eran susceptibles —impétigo—, pero aún sin el desenlace —GPE—).

Para el proceso de explicación de tamaño de muestra y asignación se va a desarrollar un ejemplo de un estudio analítico longitudinal (prospectivo), en el que la asignación es experimental al azar (experimento controlado y aleatorizado [ECA]). Muchos de los principios que se explicarán se aplican total o parcialmente a los otros diseños analíticos (experimentos controlados no aleatorizados, cohortes, casos y controles).

Muestreo, asignación y tamaño de muestra

Si retomamos el ejemplo de penicilina V *vs.* el descostrado en jóvenes con impétigo por EBHGA, es claro que la muestra (la cohorte) va a ser dividida en al menos dos grupos: 1) unos sujetos que reciben penicilina V, y 2) otros que solamente reciben medidas locales. En ambos grupos se determinará la incidencia de GPE.

El proceso de ubicar a sujetos en uno u otro grupo se denomina ***asignación***. En la tabla 5.1 se ilustra cómo la cohorte que compone la muestra se asigna a dos grupos, así como la exposición al factor de interés (penicilina V) y la exposición al factor de control (sin penicilina V); al hacerle seguimiento en el tiempo a la cohorte así incluida y asignada, se puede establecer la incidencia de daño renal en ambos grupos y se pueden comparar las frecuencias (RR, o riesgo relativo); o sea, contestar la pregunta de si existe asociación entre exponerse a la penicilina V y la disminución del riesgo de daño renal por GPE.

¿Cuántos sujetos hay que incluir en la muestra (la cohorte)? ¿Cuántos sujetos deben ser asignados a cada intervención? El problema del tamaño de muestra aquí es distinto del que se enfrentó en estudios descriptivos. Básicamente, se quiere establecer si hay alguna diferencia en la incidencia del desenlace (daño renal) ente los mismos sujetos (toda la cohorte incluida en la muestra) si todos ellos reciben penicilina V, o si todos ellos reciben placebo de penicilina y manejo local. Pero resulta imposible realizar dicho estudio. Si todos los sujetos reciben penicilina V y son seguidos durante varios meses hasta determinarse si se produce daño renal, no se puede establecer el efecto que habría surtido el hecho de no administrar penicilina V. Como en el ejemplo de la caja de fósforos, una vez evaluado un sujeto (una vez prendido un fósforo), es imposible evaluar el efecto de la otra opción (volver a encender un fósforo quemado).

Tabla 5.1 Asignación a los tratamientos y evaluación de desenlace renal

	GPE	Sin GPE	Frecuencia de daño renal (GPE)
Penicilina V	A	B	A/(A+B)
Placebo	C	D	C/(C+D)
Frecuencia de exposición	A/(A+C)	B/(B+D)	$RR = \frac{A/A+B}{C/C+D}$

Notas: Asignación de la muestra (la cohorte seleccionada y ensamblada a partir de la población elegible y accesible) a cada una de las exposiciones: Penicilina V y placebo de penicilina y seguimiento para comparar las incidencias de daño renal. RR = riesgo relativo.

La única forma no sesgada de estimar la frecuencia de GPE con cada una de las exposiciones (penicilina y placebo), es tomar una muestra aleatoria del grupo total incluido (de la cohorte) y exponerlo, por ejemplo, a penicilina V, y obtener un estimador de la incidencia de GPE, y, así mismo, tomar otra muestra aleatoria del mismo grupo y exponerlo (asignarlo) a no recibir penicilina y estimar la incidencia de GPE, y compararlas, tal como se indicó en la tabla 5.1. En realidad, no se estima cuántos sujetos deben ser incluidos en la muestra (en la cohorte), sino cuántos sujetos se necesita asignar a cada grupo de exposición para poder estimar si existe una diferencia entre los desenlaces entre los grupos de exposición, y cuál es su magnitud.

Para estimar cuántos sujetos deben ir en cada grupo (cada muestra aleatoria con la que se asignarán a cada grupo los sujetos de la cohorte seleccionada), es necesario establecer varias cosas:

a. ¿Qué tan grande es la diferencia en la frecuencia o la magnitud de los desenlaces (en el caso del ejemplo, en las incidencias de daño renal) que sería importante clínicamente, y, por tanto, valdría la pena ser capaz de demostrar? Si la diferencia que se espera hallar es muy grande, puede hacerse evidente estudiando a pocos sujetos. Dicho de otro modo (y retomando los conceptos de tamaño de muestra, ya enunciados cuando se habló de estudios descriptivos), la estimación de la diferencia no necesita ser muy precisa; pero, por el contrario, si la diferencia esperada (o relevante) es pequeña, se requiere una estimación más precisa y un mayor número de sujetos asignados a cada grupo; o sea, un mayor tamaño de muestra.
b. ¿Cuál es la probabilidad de cometer un *error tipo I* a la que el investigador se puede arriesgar? El error tipo I consiste en afirmar equivocadamente que la diferencia hallada en desenlaces entre los grupos es verdadera, cuando, en realidad, no existe, y es producto una estimación imprecisa. Mientras más pequeña sea la magnitud del error tipo I que se está dispuesto a tolerar, mayor es el tamaño de muestra. A la probabilidad de cometer un error tipo I se la denomina *p* (probabilidad de alfa), y es la famosa "p" que se reporta en los estudios para explicar la significación estadística de una diferencia.
c. ¿Cuál es la probabilidad de cometer un *error tipo II* que el investigador puede tolerar? El error tipo II ocurre cuando se afirma que una diferencia no es estadísticamente significativa (la p es elevada; es decir, no significativa), pero, en realidad, hay diferencia en los efectos de las exposiciones comparadas. La probabilidad de cometer un error tipo II es la p (probabilidad de beta), y mientras más pequeña sea p , un mayor número de sujetos deberá ser asignado a cada exposición. Al complemento de p se lo conoce como *poder de un estudio*. Si la p es 0,1 (10 %), el poder será 0,9 (90 %). Dicho de otro modo, la probabilidad que tiene un estudio con dicho poder, de hallar una diferencia de un tamaño dado y para un nivel específico de error tipo I, en caso de que esta verdaderamente exista, es del 90 %.
d. ¿Cuál es la variabilidad del desenlace en la población fuente? Mientras mayor sea la variabilidad, más sujetos son necesarios para encontrar una diferencia, si esta existe.

EL PROCESO DE MUESTREO

El muestreo puede ser, de acuerdo con sus características básicas, de dos tipos: 1) probabilístico y 2) no probabilístico, o ***de conveniencia***, según el empleo de mecanismos objetivos o subjetivos, la presencia o la ausencia del criterio del investigador y el conocimiento ***a priori*** de la probabilidad de que cada unidad de la población forme parte de la muestra. Estos métodos se describen a continuación, junto con las principales técnicas de selección, que se emplean una vez se ha definido el tamaño de la muestra en cada caso.

Muestreo probabilístico

En este tipo de muestreo, cada miembro de la población tiene una probabilidad conocida y superior a cero de ser incluido en la muestra, y las unidades muestrales hacen parte de la muestra independientemente del criterio o del gusto del investigador. Este método se considera el mejor para asegurar la validez de cualquier inferencia o generalización hecha con base en los resultados obtenidos a partir de la muestra, por las razones expuestas en los apartes anteriores.

Técnicas de selección en muestreo probabilístico

Muestreo aleatorio simple

Es la técnica más sencilla y directa de muestreo probabilístico. En ella, cada sujeto o cada unidad tienen una probabilidad igual y conocida de ser seleccionados. Supone que no existen razones para considerar que un conjunto específico de la población debe ser seleccionado con prioridad por sobre otro. Debido a tan importante característica, las muestras aleatorias simples son consideradas libres de sesgos.

La forma típica de seleccionar una muestra en este caso implica el uso de un mecanismo de probabilidad bien diseñado, el cual garantice que la probabilidad de un elemento con una característica particular de ser incluido en la muestra sea proporcional a la frecuencia de dicha característica en la población fuente (***espacio de muestreo***). **El primer paso para ello** consiste en enumerar todos los elementos de la población de manera unívoca y exacta, etapa que previamente requiere haber identificado de manera correcta todas las unidades sobre el marco muestral (espacio de muestreo).

Finalmente, y a través de algún procedimiento de sorteo (tablas de números aleatorios, números aleatorios generados con un dispositivo móvil [celular, iPad o tableta] o una lista aleatoria generada por computador, etc.), se elige a tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Un error que no pocas veces se comete durante el sorteo de las unidades que conformarán la muestra es descartar una unidad seleccionada porque no se la considera adecuada. Tal consideración destruye el supuesto probabilístico de la muestra y lleva a la introducción de sesgos.

Muestreo sistemático

El conjunto de elementos que constituye la población fuente para obtener la muestra se ordena secuencialmente.

EJEMPLO

La población de interés está constituida por sujetos que consulten a un servicio de urgencias y reciban una impresión diagnóstica de neumonía. Los casos, entonces, se ordenan, por ejemplo, según la secuencia de llegada: sujeto 1, sujeto 2, sujeto 3… sujeto *n*.

Luego, mediante un proceso aleatorio, se escoge un número que corresponde a un intervalo de sujetos dentro de la secuencia. La selección (***inclusión***) **de las unidades de muestreo se** hace, pues, empleando ese intervalo de medida constante.

La aplicación de este procedimiento requiere, de manera inicial, determinar el ***intervalo muestral*** **(*k*), intervalo que se calcula** dividiendo el tamaño de la población (***N***) entre el tamaño de la muestra calculada (***n***) **y aproxi**mando el resultado al número entero más próximo. A continuación, se debe seleccionar aleatoriamente una de las primeras ***k*** **unidades del marco muestral, la cual se denomina** *unidad de arranque* **del proceso de selección.** Finalmente, para seleccionar las demás unidades que formarán la muestra, se le suman a la posición de la unidad de arranque los valores k, $2k$, $3k$,..., **etc., hasta completar el tamaño de** la muestra requerido, y se localizan las unidades del marco muestral correspondientes a los valores obtenidos.

EJEMPLO

Siguiendo con el ejemplo de la población fuente de sujetos con neumonía, el investigador necesita seleccionar a 10 sujetos (*n*) de una población fuente de sujetos con neumonía de 120 pacientes. *K* corresponde a 120 / 10 = 12. El número aleatorio que usará como intervalo muestral debe estar entre 1 y 12. Por ello, decide lanzar 2 dados y obtiene el número 9 (ese número es producto de un proceso —experimento— aleatorio). Escoge, entonces, como su primer elemento muestral (unidad de arranque) al paciente número 9. A partir del sujeto número 9, incluye a cada noveno paciente, hasta completar el tamaño de muestra predeterminado. Ocasionalmente, la cantidad de sujetos disponible en realidad es menor que el *N* esperado. En esos casos, puede ocurrir que llegue al final del listado de pacientes sin completar el tamaño de muestra. Entonces, continúa la secuencia a partir del primer sujeto en adelante. Si solo hay 105 sujetos y el siguiente elemento muestral para incluir es el sujeto 107, regresa al sujeto 1 (106), e incluye al sujeto 2 (107).

Este es un sistema verdaderamente aleatorio y muy práctico, pues no requiere la enumeración previa de las unidades del marco muestral; basta con que estén físicamente ordenadas en un orden ascendente o descendente, para permitir el conteo.

La importancia de sortear la unidad de arranque en el proceso de selección se debe a que al fijar un número de arranque, el resto de las unidades que formarán la muestra quedan seleccionadas de forma simultánea; como si se sorteara un grupo completo.

Cabe anotar que el orden en el cual se dispongan las unidades es muy importante, debido a la homogeneidad o la heterogeneidad entre los grupos y dentro de ellos. Por tanto, la ordenación que producirá eficiencia óptima debe ser la que garantice que las unidades seleccionadas sean muy diferentes entre sí. Básicamente, el ordenamiento debe obedecer a algún proceso no determinístico, y, sobre todo, no regulado por el investigador que obtiene la muestra.

Muestreo estratificado

En algunas ocasiones, la población que se quiere estudiar se halla compuesta por subgrupos bien definidos, que pueden ser identificados con anterioridad al proceso de selección de la muestra, y, adicionalmente, la característica de interés en el estudio varía ampliamente entre esos subgrupos definidos; es decir, existe gran heterogeneidad entre ellos, en tanto que dentro de cada subgrupo la variabilidad es poca; o sea, existe homogeneidad dentro del subgrupo. En tales condiciones, pasar por alto la heterogeneidad que presenta la característica de interés sobre el marco muestral puede conllevar que las estimaciones obtenidas a partir de la muestra difieran considerablemente de los verdaderos valores de la población.

Una manera de controlar el efecto de la heterogeneidad sobre los estimadores consiste en agrupar los elementos de la población en subgrupos homogéneos dentro de sí y seleccionar una muestra de cada uno de los subgrupos. En dicho procedimiento, los distintos subgrupos reciben el nombre de ***estratos*, y el procedimiento de** agrupar o formar los estratos recibe el nombre de ***estratificación*.**

La estratificación puede, en algunas ocasiones, obedecer a características de tipo espacial, tales como el sitio de vivienda, lo cual puede identificar una clase económica determinada que sigue ciertos comportamientos sociales, políticos, económicos y culturales. En otros casos, la estratificación se basa en una o más variables sobre las cuales se tiene suficiente conocimiento. Estas variables pueden ser ***cualitativas*** **(sexo, estado civil, creencias religiosas,** identidad política, etc.) o ***cuantitativas*** (edad, peso, ingresos, etc.).

¿Cómo establecer el número de unidades que serán seleccionadas en cada subgrupo? Esta pregunta debe resolverse después de conformar o identificar los estratos, y dependerá del conocimiento que se tenga de cada uno de ellos. Hay diferentes formas de hacer la selección o la *afijación* **de las unidades dentro de los estratos.** Las más conocidas y utilizadas son:

a. **Afijación igual:** **Consiste en distribuir equitativamente** el tamaño de la muestra en cada uno de los estratos de la población; requiere conocer el número de estratos existentes. Así, por ejemplo, si existen 3 estratos identificados y se requiere un tamaño de muestra de 30 unidades, se procede a seleccionar de manera independiente 10 unidades o elementos de cada estrato, mediante las técnicas de muestreo aleatorio simple o de muestreo sistemático.
b. **Afijación proporcional:** **La distribución del tamaño** de muestra se hace porcentualmente, según el tamaño de cada estrato dentro de la población; por tanto, requiere conocer el número de estratos y el tamaño porcentual de cada uno en la población. Ejemplo: Se tienen 3 estratos, que corresponden al 50 %, al 30 % y al 20 %, respectivamente, de la población. Un tamaño de muestra de 30 unidades sería recolectado de la siguiente manera: 15 unidades en el estrato 1, 9 unidades en el estrato 2, y las restantes 6 unidades, en el estrato 3.
c. **Afijación óptima:** **Requiere conocer, además del** número de estratos y del tamaño porcentual de estos en la población, la desviación estándar poblacional en cada estrato. Consiste en balancear la variabilidad dentro de los estratos con su tamaño. Generalmente, es la afijación que producen los estimadores con la menor variabilidad.

Muestreo por conglomerados

Los conglomerados son agrupaciones de elementos que existen naturalmente (no los define el investigador) por razones económicas, biológicas, sociales, etc. (el hogar, un curso de estudiantes de un colegio, una empresa, un municipio o una ciudad). Una muestra por conglomerados es una muestra aleatoria simple, en la cual las unidades muestreadas son colecciones o grupos de unidades del marco muestral.

El propósito del muestreo por conglomerados es disminuir los costos de recolección de la información; en especial, los que corresponden al desplazamiento entre las unidades muestrales. Sin embargo, esta "reducción de costos" conlleva una pérdida de eficiencia de los estimadores que se quieren obtener; o sea, la varianza de estos puede ser mayor, comparada con el uso de los anteriores métodos de muestreo. Cuando se selecciona la técnica de conglomerados como método para obtener una muestra, se deben cumplir ciertas reglas, como las siguientes:

a. **Los conglomerados deben ser definidos de** tal manera que todos los elementos de la población pertenezcan a uno, y solamente a uno, de los conglomerados.
b. **Se debe tener una aproximación del número** de elementos que conforman cada conglomerado.
c. **Con el fin de lograr una disminución en los** costos de recolección, los conglomerados deben ser de tamaño pequeño. Es importante notar que los conglomerados no deben estar definidos de igual manera para toda la población, y pueden variar en el tamaño o en el número de unidades que los conforman. Por ejemplo, en un mismo estudio un conglomerado puede ser una casa, un edificio o una manzana. Aunque la igualdad de tamaño para los conglomerados es deseable, no es un requisito indispensable, y en la práctica es difícil de alcanzar.

La homogeneidad y la heterogeneidad entre conglomerados y dentro de ellos son muy importantes en este diseño muestral, pues tales características repercuten en la eficiencia estadística de los estimadores calculados. Es decir, la varianza de los estimadores obtenidos mediante la aplicación de esta técnica de selección muestral es mayor que la obtenida mediante muestreo aleatorio simple o muestreo estratificado, pues la similitud entre las unidades de un mismo conglomerado hace que la mayoría de la información recolectada dentro de un mismo conglomerado sea redundante.

EJEMPLO

En una encuesta en la que se definen como conglomerado las viviendas, al indagar por el canal de televisión más visto se puede obtener, más o menos, la misma respuesta por la mayoría de las personas de la vivienda encuestada.

Muestreo polietápico

La práctica del muestreo en contextos reales combina a menudo más de un método, con el fin de minimizar costos o maximizar los resultados. De manera general, dicho procedimiento consiste en la extracción de muestras cada vez menores, obtenidas de muestras más grandes en etapas sucesivas, hasta tener representatividad de las unidades de la población que se quiere investigar.

EJEMPLO

Se quiere analizar el tiempo promedio de permanencia de los pacientes en las unidades de cuidado intensivo (uci) **del país. La necesidad de emplear una** muestra es evidente, pues el recuento de todas las unidades de la población resultaría en costos excesivos y dificultades logísticas inimaginables.

Puede, entonces, tomarse una muestra aleatoria simple de pacientes en todo el país; sin embargo, algunos aspectos logísticos (costos, levantamiento del marco muestral, etc.) presentarían grandes dificultades operativas que no permitirían el desarrollo de tal proyecto. Debe considerarse, además, la posibilidad de que en un país haya subgrupos de UCI (las que tienen tecnología de punta, las unidades generales, las unidades médico-quirúrgicas, las unidades coronarias, las de menos de 5 camas, las más de 30 camas, etc.), entre las que puede haber diferencia en muchas variables (gravedad de los pacientes, días-estancia, etc.). Dicha dificultad hace necesario pensar en un método de muestreo diferente.

Si bien la opción de un muestreo estratificado contribuye con el control de la variabilidad en los subgrupos mencionados, tendría similares limitaciones operativas que las de un muestreo aleatorio simple. Por tanto, lo más recomendable es adoptar un procedimiento de selección que, por una parte, considere los estratos, e, igualmente, considere unidades de muestreo a las **UCI (conglomerados),** y a los pacientes, unidades finales de análisis; esto sugiere la adopción un ***diseño muestral polietápico*** (de tres etapas), donde las unidades primarias de análisis son los estratos; las unidades secundarias, las UCI, y las unidades finales, los pacientes.

Muestreo no probabilístico

En algunas ocasiones, el muestreo probabilístico resulta en exceso costoso e imposible de considerar, o, simplemente, su realización no es factible.

Se recurre, entonces, a métodos no probabilísticos, aun a sabiendas de que no sirven para hacer generalizaciones, pues, según como ya se explicó, la única forma de que el indicador calculado en una muestra sea un estimador no sesgado del parámetro de la población fuente es que la inclusión de los elementos en la muestra sea el resultado de un proceso aleatorio (probabilística). En general, en este tipo de muestreo, las unidades se seleccionan por conveniencia, de manera secuencial, siguiendo determinados criterios subjetivos, o, simplemente, según su disponibilidad.

EJEMPLO

Un investigador quiere evaluar un hipoglucemiante nuevo, para lo cual requiere a pacientes con diabetes tipo 2 en número suficiente para tener un grupo de intervención, y uno de control. Dada la relativa facilidad para obtenerlos en poco tiempo y en una sola institución, puede hacer muestreo probabilístico.
El mismo investigador quiere evaluar una intervención farmacológica en personas con una forma rara de dislipidemia familiar. Es posible que le resulte difícil conseguir el tamaño de la muestra, y deberá invertir mucho tiempo y esfuerzo en obtener a los sujetos. Incluso, tendrá que recurrir a varias instituciones en la misma ciudad, y, posiblemente, en varias ciudades. En este caso, permitir al azar que decida si ingresa el paciente conseguido con tanto esfuerzo no tiene mucho sustento práctico.

Adicionalmente, el muestreo probabilístico puede no ser la mejor metodología cuando se pretende describir un fenómeno poco conocido. O cuando resulta arriesgado incluir o excluir para la investigación a determinados miembros de la población (líderes comunitarios, líderes naturales, etc.) o cuando la exclusión de ciertos miembros de un grupo puede ser interpretado como algo inoportuno y ofensivo. Tampoco, cuando se deben crear constructos sociales (para su verificación mediante diseños posteriores más rigurosos). O, finalmente, cuando ciertas consideraciones de carácter logístico o ético desaconsejan su realización: por ejemplo, puede ser más fácil y barato incluir a determinados miembros de la población, o no se prevé que se justifiquen los costos o los esfuerzos adicionales que representaría. En los mencionados casos, los criterios de selección para investigar instituciones, regiones o poblaciones especiales son distintos de los necesarios para obtener una muestra probabilística representativa.

Algunos métodos de muestreo ***no probabilísticos*** **son los siguientes:**

Métodos de muestreo no probabilístico

Muestreo por conveniencia

En este método de muestreo, las unidades son seleccionadas porque resultan convenientes para el investigador, o porque se piensa que pueden representar, más o menos, a la población de interés. Ejemplos de tal tipo de muestreo son las muestras obtenidas en instituciones de salud, en consulta externa, del grupo de pacientes de un médico, de voluntarios o de los grupos autoseleccionados (o disponibles).

Muestreo por juicios

En esta técnica, los elementos de la muestra son elegidos a criterio del investigador, sobre la base de lo que él cree que dichos elementos seleccionados pueden aportar al estudio. Este tipo de muestreo se emplea, sobre todo, en investigación cualitativa.

Muestreo por cuotas

El investigador utiliza en esta situación el conocimiento previo que tiene sobre la población, con el fin de hacer "más representativa" la muestra. Se suelen identificar diferentes estratos relevantes de la población, y en la selección de la muestra se toma en cuenta la proporción de cada uno de ellos. En este tipo de muestreo, se fijan unas *cuotas*, **consistentes en un número de individuos** que reúnen unas determinadas condiciones; por ejemplo: 20 individuos de 25-40 años, de sexo femenino y residentes en La Candelaria, Bogotá. Una vez determinada la cuota, se elige a los primeros que se encuentren y que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión y de mercadeo, así como en algunos diseños epidemiológicos en los que se necesita fracciones de muestreo altas en ciertos subgrupos. Tal como ya se describió, el muestreo no probabilístico por cuotas tiene su contraparte probabilística en el muestreo por estratos, en el cual se puede buscar que la cantidad de elementos de cada estrato sea la suficiente para hacer estimaciones específicas (cambiar las fracciones de muestreo para que la afijación sea igual), y no que el tamaño de los estratos sea proporcional a su frecuencia en la población fuente (***no afijación proporcional***).

EJEMPLO

En un estudio, se quiere describir la composición de la leche materna "prematura" examinando muestras de leche materna provenientes de madres con bebés prematuros vivos, de diferentes edades gestacionales. Debido a que la frecuencia de nacimientos prematuros cercanos al término es mucho mayor que la de los prematuros muy inmaduros, al reclutar, por ejemplo, a 60 diadas madre-hijo con nacimiento prematuro, la gran mayoría de las unidades muestrales tendrán edades gestacionales cercanas al término (34-36 semanas de edad gestacional), y en muy pocas diadas el recién nacido será extremadamente prematuro (menos de 30 semanas de edad gestacional). Para lograr suficientes sujetos de todas las edades gestacionales de interés, se pueden utilizar 3 "cuotas" de 20 sujetos: 20 entre 34 y 36 semanas, otros 20 entre 30 y 33 semanas y otros 20 de menos de 30 semanas. Se recluta secuencialmente a las diadas. Cuando se alcanza el tamaño de una cuota (por ejemplo, las 20 diadas de 34-36 semanas), se suspende el reclutamiento de ese grupo de edad gestacional y se continúa en las otras edades hasta alcanzar los tamaños de muestra de cada cuota.

Muestreo de bola de nieve

Este tipo de muestreo consiste en pedir a cada uno de los sujetos que cumplen los criterios de admisión que sugiera o presente a otros posibles candidatos para ser incluidos en la muestra. Este es el método recomendado para el estudio de casos de interés especial, difíciles de identificar, tales como personas con enfermedades muy raras, drogadictos, personas que han sufrido agresión sexual, miembros de una secta, etc. Es una técnica apropiada para la investigación cualitativa.

RECAPITULACIÓN

A continuación se ofrece una breve síntesis de los principales puntos revisados.

Preguntas descriptivas

En el caso de las preguntas descriptivas, se busca estimar uno o varios parámetros de una población blanco o diana, a través de una muestra idealmente probabilística. La muestra (muestreo) se toma de una población elegible (población blanco + criterios de inclusión y de exclusión específicos), y que, además, sea accesible (una fracción de la población blanco, compuesta por sujetos elegibles y que sean accesibles al investigador; proceso de selección). La muestra es válida en la medida en la que sea una representación adecuada (probabilística, no sesgada) de la población fuente. La estimación se generaliza a la población fuente (población accesible). Si la población blanco es directamente la población accesible, la muestra estima de forma no sesgada los parámetros de dicha población blanco. Si no es así, lo que se hace es extrapolar los estimativos de una población accesible a toda la población blanco. El tamaño de muestra depende de la precisión deseada (anchura del intervalo de confianza [IC] en torno a la estimación específica) del grado de confianza (por ejemplo, del 95 %) y de la variabilidad de la característica de interés en la población fuente. A mayor precisión (intervalo más estrecho, o nivel de confianza más elevado), mayor número de sujetos. A menor variabilidad en la población, menor número de sujetos. El concepto de población de referencia es poco relevante en este contexto.

Preguntas analíticas

No se busca estimar un parámetro poblacional, sino determinar si los sujetos que pertenecen a una población abstracta (población de referencia), por el hecho de hallarse expuestos a un factor determinado, pasan a formar parte de otra población con un riesgo diferente de un desenlace de interés. Se identifica (selección) una población elegible y accesible (tiempo, lugar y persona, criterios de inclusión y de exclusión) que se constituya en un buen modelo de la hipótesis que se quiere evaluar (población de prueba). En dicha población (que es, básicamente, una cohorte) se toman muestras (***aleatorias*** = asignación experimental; ***no aleatorias*** = asignación observacional) de sujetos con y sin la exposición de interés. El ***muestreo*** es la asignación de sujetos de la cohorte a las diferentes exposiciones. El tamaño de muestra (total de sujetos asignados a todos los grupos, y número necesario en cada grupo) depende de la diferencia en desenlaces que se espere encontrar, de los niveles de los errores tipo I y II que se pueda tolerar y de la variabilidad de la presentación del desenlace en la población elegible y accesible.

La validez interna es el grado en el cual la comparación de las dos muestras tomadas de la población elegible y accesible es no sesgada y precisa (tanto en la asignación como en la evaluación), y la validez externa se traduce en la confianza que se puede tener al extrapolar los resultados a sujetos distintos de los estudiados, que pertenecen a diferentes subpoblaciones de la población de referencia. El concepto de población blanco, en principio, no es relevante, pues no se hacen estimaciones específicas, sino inferencias inductivas (extrapolaciones a integrantes de la población de referencia, en vez de generalizaciones a una población fuente concreta). El concepto de población blanco solo se aplica en los casos especiales donde la predicción o la explicación son de interés únicamente para una población concreta (estimación de un efecto en una población), y no se busca extrapolar (modelar una asociación que sea relevante en una población abstracta, de referencia).

Proceso de muestreo

Los procesos de muestreo descritos son de relevancia, fundamentalmente, para estimaciones de parámetros poblacionales a partir del examen de muestras; es decir, para preguntas descriptivas y poblaciones blanco. Hay ocasiones en las que se formulan preguntas analíticas en poblaciones blanco, y no de referencia (por ejemplo, experimentos pragmáticos), en los cuales los procesos de muestreo descritos también son relevantes. En el caso de estudios analíticos que buscan evaluar hipótesis sobre poblaciones de referencia (abstractos), el proceso de selección de sujetos para conformar la muestra de prueba no busca la representatividad estadística de una población blanco (definida), sino la representatividad conceptual de una población de referencia.

Los tipos de muestreo se dividen en ***probabilísticos*** y ***no probabilísticos***. Solo los muestreos probabilísticos aseguran remover los sesgos de selección producidos durante el muestreo, y permiten hacer una estimación no sesgada (generalización) de los parámetros poblacionales de la población fuente (el conjunto del cual, de forma directa, se selecciona la muestra).

Los muestreos probabilísticos se dividen en ***aleatorios simples*** y en ***aleatorios con otros procesos***: secuenciales, estratificados, por conglomerados y polietápicos.

Los muestreos no probabilísticos no permiten hacer estimaciones no sesgadas de parámetros de una población fuente, pero sí permiten conducir una investigación cualitativa (generación de hipótesis) y pueden ser la única alternativa factible para hacer investigación cuantitativa en circunstancias especiales.

Lecturas sugeridas

Celis de la Rosa AJ. Bioestadística. 2ª. ed. México: Manual Moderno; 2008.

Cochran WG. Sampling techniques. 3rd ed. Hoboken: John Wiley and Sons; 1977.

Dawson-Saunders T. Bioestadística médica. 2ª. ed. México: Manual Moderno; 1999.

Greenland S. Target population. En: Armitage P, Colton T. Encyclopedia of biostatistics. Chichester; New York: J. Wiley; 1998.

Greenland S. Validity and generalizability in epidemiologic studies. En: Armitage P, Colton T. Encyclopedia of biostatistics. Chichester; New York: J. Wiley; 1998.

Hulley SCS. Designing clinical research. An epidemiological approach. Baltimore: Williams and Wilkins; 1999.

Kish L. Survey sampling. New York: Wiley; 1965.

Lwanga SK, Lemeshow S. Sample size determination in health studies: A practical Manual. Ginebra: OMS; 1991.

Milton JS. Estadística para biología y ciencias de la salud. 3ª. ed. Madrid: McGraw-Hill; 2004.

Pagano M, Gauverau K. Principles of biostatistics. Duxbury: Duxbury Press; 2000.

Rothman KJ, Greenland S, Lash T. Validity in epidemiologic studies. En: Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008. p. 758.

Sedgwick P. Randomised controlled trials: the importance of sample size. BMJ. 2015;350:h1586.

Streiner DL. Statistics commentary series: Commentary #9-Sample Size Made Easy (Power a Bit Less So). J Clin Psychopharmacol [internet]. 2015 [citado 2015 abr. 6]. Disponible en: http://journals.lww.com/psychopharmacology/Citation/publishahead/Statistics_Commentary_Sqeries__Commentary__9_Sample.99290.aspx (Última consulta 6 de abril de 2015).

CAPÍTULO

6 Cálculo del tamaño de muestra

Martín Alonso Rondón Sepúlveda

INTRODUCCIÓN

Una de las preguntas que más a menudo se hacen los investigadores al planear un estudio es: ¿cómo se determina, o cuál debe ser, el número de unidades de observación (personas, hogares, animales, objetos, etc.) que deben estudiar?[1-3].

La importancia clínica del tamaño de la muestra en investigación es enorme, aunque a menudo desconocida. Si se evalúa un medicamento en un estudio con un tamaño de muestra insuficiente, se puede llegar a la falsa conclusión de que no es útil, cuando, en realidad, el proyecto era ineficaz para encontrar una diferencia significativa, tan solo por su bajo número de observaciones.

También puede ocurrir que un experimento clínico demuestre que un procedimiento es exitoso y mejor que la intervención control, pero lo logra con un número significativamente mayor que el necesario, lo cual significa que hubo más costos, más riesgos y más trabajo que el que hacía falta, con todas sus consecuencias éticas, logísticas y económicas fáciles de imaginar.

Infortunadamente, la pregunta de cuántas observaciones son necesarias no tiene una respuesta única, pues su resultado depende de múltiples factores, como cuál es la población de estudio, cuál es el objetivo de interés de la investigación, cuál es el número de grupos que se va a estudiar, cuál es la homogeneidad en los resultados de la variable de interés, cuál es la precisión y cuál es la confiabilidad con la que se quiere hacer tal estimación, por mencionar solo algunos.

Contar con la mencionada información permite hacer estimaciones más reales sobre el tamaño de muestra que se necesita estudiar. Sin embargo, uno de los problemas más comunes al respecto es que no siempre se tiene certeza de esos datos, o, simplemente, no se dispone de ellos, lo cual conlleva incertidumbre en el momento de hacer su cálculo. Siempre se espera contar un tamaño de muestra óptimo, de tal forma que se puedan hacer inferencias a la población con la mayor certidumbre posible; ojalá, tratando de minimizar el error en la estimación y asegurando un alto nivel de confianza.

Hacer censos o tomar tamaños de muestra muy grandes no garantiza resultados más confiables, así como tamaños de muestra pequeños no significan que el estudio no tiene validez. Trabajar con una muestra de la población tiene la ventaja de que es más rápido y económico, y en caso de lograr una adecuada representatividad de la población permite que los resultados precisos puedan ser extrapolados a la población

blanco y que se pueda estimar de una forma más razonable una situación determinada en dicha población[4].

Por lo anterior, es importante determinar de forma específica, para cada estudio, cuál debe ser el número mínimo de unidades de observación que se deben considerar, de forma que al final se pueda tener tranquilidad en que los resultados son válidos y confiables.

El objetivo principal del presente capítulo es introducir al tema del cálculo del tamaño de muestra en investigación, y busca presentar cómo algunos conceptos básicos, como la pregunta de investigación, los errores, la variabilidad, el tamaño del efecto, etc., deben considerarse dentro del cálculo del tamaño de muestra para un estudio de investigación. Se presentan algunas fórmulas y ejemplos para los casos más comunes, considerando, especialmente, el tipo de desenlace (dicotómico o continuo) en los diseños más habituales (estudios de prevalencia, de cohorte, de casos y controles y experimentales).

ASPECTOS IMPORTANTES

Para desarrollar un proyecto de investigación, es importante considerar una gran variedad de aspectos que aseguren la calidad de todas y cada una de las etapas de realización del estudio[2,3,5-8]. Hacer estudios con la mayor rigurosidad no necesariamente asegura que los resultados hallados sean los reales, pero sí le da al investigador mayor tranquilidad y confianza en ellos; por eso, cualquier descuido en el proceso puede ocasionar deficiencias para responder la pregunta de investigación.

Uno de mencionados aspectos es el concerniente al número de unidades de observación (ya sean individuos o elementos) necesarios para garantizar la validez y la solidez en los resultados[6]. Hacer un análisis de toda la población parecería ser lo ideal, aunque ello resulta imposible en la práctica, debido, principalmente, y por mencionar solo algunos de los aspectos, a la dificultad para acceder a todos los sujetos de la población de estudio, al gran consumo de recursos (especialmente, económicos) y a los tiempos limitados para la realización del estudio; y, en algunos casos, porque, simplemente, no es factible, pues se debe destruir la unidad de análisis. Es necesario, entonces, seleccionar solo una parte de la población para su estudio; o sea, ***seleccionar una muestra.***

El problema se centra, pues, en determinar cuál debe ser el tamaño mínimo necesario de unidades de observación que se deben seleccionar en la muestra del caso, de tal forma que les permita a los investigadores encontrar resultados válidos, confiables y, ojalá, precisos. Tamaños de muestra demasiado pequeños pueden no representar lo que en realidad está sucediendo en la población (en especial, cuando la población de estudio es bastante heterogénea); por lo tanto, se puede tener bastante incertidumbre en los resultados, y, probablemente, no podremos ser precisos al hacer la estimación de una o más medidas de interés (promedios, proporciones, tasas, etc.). En el otro extremo, considerar tamaños de muestra excesivamente grandes, además de hacer que la realización del estudio se torne más difícil y costosa, puede conducir a pérdidas de tiempo y de recursos, y en algunos casos, a reducir la calidad del estudio[9].

En ese orden de ideas, el cálculo del tamaño de muestra se convierte en un aspecto relevante a la hora de hacer un estudio o una investigación. Muchas veces, los investigadores o los grupos de investigación deciden, por motivos "prácticos", el número máximo de participantes para sus estudios, y solo recurren a asesoría estadística cuando los comités de ética o los editores de revista piden que esta se encuentre por completo justificada. Lo ideal es disponer de un grupo interdisciplinario en el momento de desarrollar el protocolo del estudio, de manera que aspectos como el tamaño de muestra sean discutidos y estimados desde dicho momento. En otros casos, se preguntan si el número de unidades de observación con el que cuentan es suficiente para poder encontrar diferencias estadísticamente significativas (***poder***), o, simplemente, quieren saber cuál es el tamaño de muestra mínimo para su estudio, con el fin de evaluar su factibilidad.

Tomar tamaños de muestra demasiado grandes cuando no se necesitan o tomar tamaños de muestra insuficientes hace que muchas veces se hagan cuestionables la validez de los resultados y los aspectos éticos del estudio. Siempre debe buscarse el tamaño de muestra óptimo.

A continuación se describirán algunos conceptos básicos que se deben tomar en cuenta en el cálculo del tamaño de muestra para un estudio de investigación.

Pregunta de investigación, hipótesis de interés y diseño

"La pregunta de investigación es el elemento esencial del cual parte el estudio de interés"[2]; y aunque en muchos casos las preguntas son similares entre investigaciones, no siempre las metodologías empleadas en otros estudios son aplicables, incluidas las del cálculo del tamaño de muestra. Ello se evidencia más claramente cuando las características de la población son visiblemente distintas.

Por lo general, las estimaciones del tamaño de muestra se basan en las hipótesis estadísticas de interés en las que se especifican los parámetros para estimar. El investigador debe cuestionarse por anticipado si la información que va a recolectar es suficiente para probar su afirmación.

Por ejemplo, supongamos que se lleva a cabo un experimento clínico para comparar un nuevo medicamento (B) con uno ya existente (A), en un grupo de pacientes hipertensos, respecto a sus niveles promedio de presión arterial sistólica. Para ello, el investigador selecciona un grupo de pacientes al azar y los divide en dos grupos de igual tamaño. Uno de los primeros supuestos que debe probar es que los dos grupos son comparables; o sea, que las características de base de los dos grupos son similares entre sí, y que en caso de encontrar diferencias, estas se deben, específicamente, al efecto que presentan los medicamentos sobre el desenlace de interés (en este caso, reducir los niveles promedio de presión arterial).

Si el interés de la investigación es hacer estimaciones sobre una o más características de la población, se debe estar seguro de que dichas estimaciones tienen la confiabilidad y la precisión apropiadas para cada una de tales características.

Si el objetivo es evaluar diferencias o determinar la presencia de algún efecto (debido a un método o un tratamiento), se debe estar seguro de que las conclusiones derivadas sean producto de una variación propia del método o del tratamiento, y no simple producto del azar.

Según lo expuesto, es importante tener claridad sobre la direccionalidad de las hipótesis, pues hay una gran diferencia entre un tamaño de muestra para probar que un tratamiento es mejor que otro y decir que los dos tratamientos son diferentes. Desconocer la direccionalidad implica mayores tamaños de muestra.

El diseño planeado para la recolección de los datos es otro factor determinante en el momento de hacer el cálculo[2,3], y aunque esté directamente ligado a la pregunta de investigación, el tamaño de muestra puede diferir según la estrategia de diseño empleada; por ejemplo, el tamaño requerido es muy diferente si la pregunta va a ser resuelta usando un estudio de casos y controles o un estudio de cohortes.

Parámetros

Los parámetros son valores de interés obtenidos a partir de la población[10-12]. Específicamente, *parámetro* se refiere a una medida o un valor derivados de una variable, y cuyo fin es caracterizar o representar a la población; esto quiere decir que se busca, a partir de un valor, resumir aspectos, situaciones o eventos de dicha población.

Estadísticos

Los ***estadísticos*** son los valores de interés obtenidos a partir de una muestra, y son, por ello, una estimación de los parámetros de la población[10-12]; se refieren a medidas de resumen de una variable aleatoria, pero su fin principal consiste en estimar los parámetros de interés a través de los resultados de muestras aleatorias; o sea, estimar situaciones o eventos que están pasando realmente en la población, con base en los resultados de una muestra.

Error de muestreo

El ***error de muestreo*** se define como la desviación de la muestra aleatoria seleccionada respecto a las verdaderas características observadas en la población. En otras palabras, es la diferencia entre el estadístico y el parámetro de interés; es decir, entre el valor real y la estimación hecha por la muestra; por ello, cuanto más cercana al parámetro sea la estimación, menor será su error muestral. Se le considera una medida de la variabilidad de las estimaciones[2,10-12]. A mayor tamaño de muestra se espera que el error de muestreo sea menor.

Variabilidad

Este es, quizás, uno de los aspectos clave en el cálculo del tamaño de muestra. La varia-

bilidad observada en el desenlace de interés influye directamente en las estimaciones del tamaño de muestra, pues cuando las unidades de observación de una determinada población son homogéneas, implícitamente se dice que hay ***poca variabilidad*** en sus resultados, y, probablemente, no se necesiten muchas observaciones para hacer una buena estimación de sus parámetros; en cambio, a medida que la variabilidad aumente, el tamaño de muestra también se irá incrementando.

EJEMPLO

En un estudio para determinar diferencias entre dos tratamientos para la hipercolesterolemia, se necesitarán más pacientes si la población incluye a personas con hipercolesterolemia familiar, y que tienen valores muy altos, pues habrá una gran variabilidad en los valores de lípidos. Si se excluye a los sujetos con hipercolesterolemia familiar (pues así habrá más homogeneidad en los valores basales), se requerirá un tamaño de muestra menor.

Se recomienda, por lo anterior, tener conocimiento ***a priori*** sobre la variabilidad de la medida de interés en la población de estudio. Las poblaciones demasiado heterogéneas implican tamaños de muestra grandes, lo cual puede conducir, en algunos casos, a que el estudio no sea viable o a que no se pueda tener el suficiente poder para hallar evidencia estadística.

Uno de los retos más grandes está en determinar cuál es la variabilidad esperada en nuestro estudio. Para ello, es necesario hacer una revisión de la literatura o de resultados de trabajos publicados; incluso, en algunas situaciones se hace necesario hacer pruebas piloto para obtener información que ayude al investigador a tener algún grado de evidencia al respecto. Plantear erróneamente la variabilidad puede llevar a tamaños de muestra insuficientes.

Tamaño del efecto, y diferencia mínima esperada

Otro de los aspectos para considerar es el relacionado con el ***tamaño del efecto***, el cual tiene que ver con la magnitud del cambio sobre un desenlace de interés (por ejemplo, se puede evaluar cuál fue el cambio en la presión arterial después de aplicar o hacer una intervención). A medida que dicho cambio sea mayor, mejor será la ***potencia estadística***, lo cual indica que la nueva intervención presenta un mejor comportamiento que el tratamiento estándar, y se espera, entonces, que se detecte con mayor facilidad, lo cual significa que no se necesitaría un tamaño de muestra grande. A medida que tal cambio se vaya haciendo más pequeño, pero clínicamente importante, se necesitará mayor evidencia estadística para considerarlo mejor, lo cual se ve reflejado de forma directa en tamaños de muestra grandes[7].

EJEMPLO

Un estudio que quiera comparar un antiinflamatorio no esteroideo (AINE) con placebo en pacientes con artritis gotosa aguda necesitará a pocos pacientes para demostrar las diferencias, pues el tamaño del efecto esperado es grande. Si en otro estudio se quiere comparar un AINE nuevo con uno ya establecido en la práctica, como la diferencia que se espera entre los dos es menor, el tamaño de muestra requerido será mayor.

No obstante lo anterior, al igual que para el planteamiento de las hipótesis, este es un aspecto que necesita un alto grado de conocimiento del tema, por lo cual el investigador debe establecer ***a priori*** cuál es la magnitud esperada del efecto; en especial, cuando se están evaluado medicamentos o intervenciones, ya sea mediante juicios clínicos o por su experiencia; o cuál es el tamaño de la asociación que le interesaría ser capaz de detectar a partir de su muestra[2,3,13]. En dichos casos, se debe buscar información en la literatura de estudios similares, o, incluso, hacer también estudios piloto.

Otra alternativa que sugieren algunos autores es tomar un tamaño del efecto considerando cuál debería ser la diferencia mínima clínicamente importante. Si hubo un efecto real del tratamiento de tamaño importante, queremos que este se vea reflejado en los resultados; es decir, que al final se halle una diferencia estadísticamente significativa. Ahora, el interés principal se centra en cuando las diferencias son muy pequeñas, pues en tal caso es importante evaluar si dicha diferencia estadísticamente significativa es también clínicamente importante; por ello, es necesario decidir cuál debe ser la mínima diferencia clínicamente aceptable.

En caso de no tener ningún tipo de información, se puede recurrir a métodos que se basen

en la distribución de la variable basando los cálculos en los valores observados y generando puntos de corte de posibles tamaños del efecto. Cohen[14] sugirió el concepto de ***tamaños del efecto estandarizados de 0,2, 0,5 y 0,8*** como la representación de efectos "pequeños", "moderados" y "grandes", respectivamente.

Efecto del diseño

Una de las etapas que se deben hacer una vez calculado el tamaño de muestra es seleccionar las unidades de observación mediante una estrategia de muestreo. Se dice que el diseño de una muestra es ***difícil*** o ***complejo*** si para seleccionar las unidades de observación de la población objetivo esta se debe llevar a cabo por etapas (***muestra polietápica***). Ello significa que es muy difícil llegar directamente a la unidad de observación; o sea, no se podrá seleccionar a personas con un muestreo aleatorio simple. Si al momento de hacer la investigación se considera el uso de diseños muestrales complejos, se debe ajustar el tamaño de muestra, por lo que dicho proceso se conoce como el "efecto del diseño". Un parámetro depende del tipo de diseño establecido para seleccionar los elementos que integran la muestra. A tamaños de muestra iguales, el error es mayor si se utiliza un diseño complejo (por ejemplo, ***muestreo bietápico***) que si se usa un muestreo aleatorio simple (MAS).

Error tipo I y nivel de confianza

El error tipo I (α) se refiere a la probabilidad de error que estamos dispuestos a cometer por un resultado falso positivo[2,6-8,10-12,15]. Para saber qué error tipo I aceptar en la investigación, es necesario preguntarse hasta qué punto estamos dispuestos a aceptar resultados falsos positivos en nuestro estudio. Tal decisión depende del investigador, quien debe apoyar su decisión en la información que tenga sobre el problema. Cuál sea el nivel de error tipo I que se acepte depende de cuánta seguridad se deba tener, de acuerdo con los riesgos o los costos que implique el tratamiento.

La selección de α para las estimaciones de tamaño de muestra depende, principalmente, de las condiciones y las características del estudio, así como de las consideraciones del investigador. Aspectos como el presupuesto o el tiempo de ejecución dependen directamente de este, ya que valores muy pequeños de α implican tamaños de muestra grande, mientras valores grandes de α implican menores tamaños de muestra, y, por lo tanto, menor confiabilidad en los resultados.

No hay un valor universal para el error tipo I; sin embargo, en la mayoría de las investigaciones se utilizan más a menudo valores de α del 1 %, del 5 % y del 10 %. Cabe aclarar que no es necesario utilizar uno de estos tres, por cuanto el investigador queda en completa libertad de establecer cualquier otro valor. Lo importante es tener presente que a menor valor del error tipo I, mayor será el tamaño de muestra.

EJEMPLO

Si se planea un estudio de cuyo resultado depende que se proponga hacer una amputación, el error I debe ser el mínimo, dada la magnitud de la decisión. Si la intervención es muy costosa, será también importante que haya un nivel de error I menor.

El ***nivel de confianza*** se define, por su parte, como la probabilidad de que la estimación efectuada se ajuste a la realidad; en otras palabras, es la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro. En la investigación, corresponde a (1-α), donde α es el error tipo I fijado en el cálculo del tamaño de muestra[2,6-8,10-12,15].

El error tipo I equivale a preguntar cuánto está dispuesto a arriesgarse el investigador a que su muestra lo lleve a concluir que hay asociación cuando, en realidad, no existe. Para tomar tal decisión no solo se debe tener presente la precisión de los resultados, sino que, además, debe interrogarse sobre todos los aspectos éticos y económicos[9].

Lamentablemente, muchos investigadores basan sus conclusiones en decidir si rechazan o no la hipótesis nula, a partir de un valor arbitrario del α. Es decir, si ***a priori*** tomamos un valor α de 0,05 el investigador debe interpretar críticamente el valor de ***p*** obtenido, y no basarse tan solo en evaluar de manera absoluta si es mayor o menor que el valor de α establecido *a priori* como regla de decisión[16]. Fisher sugirió que si el valor de p es > 0,1, ciertamente no hay ninguna razón para pensar que lo planteado en la hipótesis alterna es verdad; pero si este valor se encuentra por debajo de 0,01, ya se puede

decir que se tiene una fuerte evidencia de que lo planteado en la hipótesis alterna es cierto. Lo recomendable, entonces, es manejar con cautela los valores cercanos a 0,05.

Error Tipo II y poder del estudio

El error tipo II establece la probabilidad de error que estamos dispuestos a cometer por un resultado falso negativo. La idea del uso del error tipo II, o β, en el cálculo del tamaño de muestra consiste en establecer qué tan aceptable es concluir, mediante la investigación, que no hay asociación, cuando realmente existe. En tal caso, cometer un error tipo II trae consecuencias éticas más evidentes. Al igual que el error tipo I, no hay un valor obligatorio, pero en la comunidad científica se considera que este no debería ser > 20 %; sin embargo, cuanto más pequeño sea, mayor será el tamaño de muestra requerido[2,6-8,10-12,15].

Directamente relacionado con la definición de error tipo II, está el concepto de ***poder del estudio***, que se define como 1-β. Es también conocido como la ***potencia de la prueba***, o *sensibilidad*, en el caso de pruebas diagnósticas, y señala la probabilidad de detectar como estadísticamente significativa una diferencia clínicamente o prácticamente importante de un tamaño especificado con antelación, si en realidad existe[1,8-11]. Para calcular el tamaño de muestra, si ello es necesario, el β debe fijarse de antemano, para asegurar una adecuada probabilidad de detectar cambios clínicamente importantes.

EJEMPLO

Un estudio que comparó dos antibióticos en una población total de 42 pacientes halló que la diferencia en control de la infección no tenía significación estadística. El resultado plantea la posibilidad de que no haya habido suficientes pacientes para hallar una diferencia con significación estadística.

Lo ideal es fijar ***a priori*** los valores de α y de β, de tal manera que se pueda reducir el riesgo de cometer errores o de manipular los resultados. Si se escogen valores altos de α y de β para reducir el tamaño de muestra, se puede estar influyendo directamente en la validez y la confiabilidad de los resultados. Es imprescindible seleccionar adecuadamente los valores de α y β, por cuanto ello permite tener mayor seguridad de que los resultados obtenidos son válidos, y se minimiza el riesgo de recomendar malas intervenciones (error tipo I) o de rechazar intervenciones que pueden ser beneficiosas (error de tipo II). Según argumentan Kelley, Maxwell y Rausch[17], es tan poca la atención que se les concede a los mencionados aspectos que hay la posibilidad de que una gran proporción de hallazgos o de evidencias estadísticamente significativas reportados en la literatura publicada correspondan, en realidad, a errores tipo I o sesgos de publicación.

El ***poder del estudio*** se refiere, entonces, a la capacidad que tiene el estudio para hallar una diferencia con significación estadística. Puede asimilarse al poder que se necesita para detectar bacterias en una muestra de orina. Si se examina a simple vista, el ojo no tiene capacidad para detectarlas; tampoco, si se examina la muestra con una lupa. Será necesario utilizar un microscopio, con un lente de poder suficiente para encontrar las bacterias. Cuando el observador diga que no encontró bacterias al examinar la muestra con una lupa, podrá afirmarse que su instrumento no tiene poder suficiente para encontrar bacterias, y dicho resultado será un falso negativo.

Ajuste por posibles pérdidas al seguimiento

Otro aspecto que se debe considerar es el relacionado con las pérdidas al seguimiento. En muchas ocasiones, algunos estudios quieren hacer un seguimiento a sus pacientes para evaluar cómo funciona una intervención a través del tiempo. En tales casos, es importante que al momento de terminar el estudio se tenga el tamaño de muestra suficiente para poder llevar a cabo los análisis estadísticos respectivos. Por ello, es necesario ajustar la muestra por un tamaño mucho más grande; o sea, es necesario sobreestimar el tamaño de muestra, de tal manera, que cuando finalice el seguimiento, la proporción de personas que fueron pérdidas del seguimiento no afecten la validez ni la precisión del estudio. El tamaño de muestra debe ser ajustado ***a priori*** por la proporción esperada de pérdidas, de tal manera que al finalizar se tenga, como mínimo, el tamaño de muestra calculado inicialmente.

TAMAÑO DE MUESTRA PARA DETERMINAR PARÁMETROS DE INTERÉS A TRAVÉS DE INTERVALOS DE CONFIANZA (IC)

En algunos estudios, el interés principal se centra en tratar de estimar un parámetro, bien sea una proporción, una prevalencia o un promedio de un desenlace de interés a través de una muestra. En esos casos, el cálculo del tamaño de muestra se basa, especialmente, en la precisión que se busca en torno al parámetro que se quiere estimar, con un determinado nivel de confianza. Cuanto más estrecho sea el intervalo de confianza (IC), mayor será la precisión, y el tamaño de muestra requerido aumentará. El nivel de confianza (Z_α), se obtiene al observar los valores de una distribución normal estándar. Los valores más utilizados son 1,28, 1,64 y 2,33, para valores del 90 %, del 95 % y del 99 % de confianza, respectivamente.

Tamaño de muestra para una proporción

En el caso de *variables categóricas*, la estimación del parámetro se hace a través de proporciones. En dicho escenario, los factores que se deben considerar para el cálculo del tamaño de muestra son: la precisión buscada (d), la proporción estimada (p) y el nivel de confianza (Z_α).

La fórmula para calcular el tamaño de muestra para proporciones está dada por:

$$n = \frac{(Z_\alpha^2)(p)(1-p)}{d^2}$$

Como ejemplo, supongamos que se quiere calcular un tamaño de muestra para estimar la proporción de personas con sobrepeso u obesidad en la población de Tunja. Según estudios en poblaciones similares, se espera que la proporción se encuentre en torno al 30 %. Si los investigadores quieren que dicha estimación se encuentre, más o menos, en el 5 % en torno al estimador, con un Z_α 90 %, el tamaño de muestra sería el siguiente:

$$n = \frac{(1{,}28)^2\,(0{,}3)\,(0{,}7)}{0{,}03^2} = \frac{1{,}64 \times 0{,}21}{0{,}0009}$$

$$= 382{,}67 \sim 383 \textit{ personas}$$

Por lo tanto, el tamaño de muestra mínimo que necesitaría el estudio es de 383 personas.

Tamaño de muestra para un promedio

En el caso de variables continuas, según el supuesto de distribución normal de la variable de interés, la estimación del parámetro se hace mediante los promedios. Los factores principales que se deben considerar para calcular el tamaño de muestra son: la precisión buscada (d), la variabilidad (S^2) y el nivel de confianza (Z_α).

En general, se utiliza la siguiente fórmula con el fin de calcular el tamaño de muestra para promedios:

$$n = \frac{(Z_\alpha^2) \times s^2}{d^2}$$

Como ejemplo, supongamos que se quiere calcular un tamaño de muestra para estimar la estatura promedio de una población de niños entre 2 y 5 años de edad de Bogotá. Según datos obtenidos en una muestra piloto, se espera que la desviación estándar sea de 10 cm. Si los investigadores quieren que la estimación de la estatura se encuentre más o menos 2 cm en torno al estimador, con un Zα 95 %, el tamaño de muestra sería el siguiente:

$$n = \frac{(1{,}96)^2 \times 10^2}{2^2} = \frac{3{,}8412 \times 100}{4}$$

$$= 96{,}04 \sim 97 \textit{ niños}$$

Por lo tanto, se necesitaría una muestra de 97 niños para hacer dicho estudio.

Es importante anotar que algunas fórmulas tanto para promedios como para proporciones también consideran para su cálculo el tamaño de la población y el efecto del diseño. En tales casos, se debe disponer de dicha información adicional para hacer el cálculo, y se deja a criterio del investigador seleccionar la que mejor se ajusta a su estudio. El tamaño de la población tiene un papel importante; en especial, cuando es pequeño.

TAMAÑO DE MUESTRA PARA EVALUAR PRUEBAS DE HIPÓTESIS

A diferencia de los estudios donde el interés principal radica en tratar de estimar un paráme-

tro, en el caso de dos grupos lo que se busca es establecer si hay una diferencia estadísticamente significativa entre dos promedios o entre dos proporciones. Para ello, es importante establecer cuál debe ser la diferencia que se considere clínicamente importante.

Tamaño de muestra para diferencia de proporciones

Nuevamente, por tratarse de variables categóricas, la estimación del parámetro se hace a través de proporciones en cada uno de los grupos. En el escenario planteado, los factores que se deben considerar para el cálculo del tamaño de muestra son: la proporción del evento de interés en cada uno de los grupos (p_1 y p_2), la direccionalidad de la hipótesis de interés (si es a una cola o a dos colas), el error tipo I (Z_α) y el poder ($Z_{1-\beta}$). Los valores críticos más utilizados del nivel de confianza y de poder se combinan en la **tabla 6.1** y se notan como $f(\alpha,\beta)$.

La fórmula para calcular el tamaño de muestra para proporciones está dada por:

$$n = 2\left[\frac{p_1(1-p_1)+p_2(1-p_2)}{(p_1-p_2)^2}\ f(\alpha,\beta)\right]$$

Por ejemplo, supongamos que se quiere calcular el tamaño de muestra para estimar si la proporción de hombres que fuman es mayor que la proporción de mujeres que fuman, en una población de pacientes que asisten a una central de urgencias. Según los investigadores, se espera que la proporción de hombres que fuman sea cercana al 25 %, mientras de la proporción de mujeres que fuman se espera que sea del 18 %. Con un $Z\alpha$ 95 % y un poder del 80 %, el tamaño de muestra sería el siguiente:

$$n = 2\left[\frac{0{,}25\,(0{,}75)+0{,}18\,(0{,}82)}{(0{,}25-0{,}18)^2}\right](6{,}18)$$

$$= 2\,(68{,}39)\,(6{,}18) = 845{,}26 \sim 846\ \textit{personas}$$

Por lo tanto, el tamaño de muestra mínimo que necesitaría el estudio es de 846 pacientes, que corresponden a 423 hombres y 423 mujeres.

Tamaño de muestra para diferencia de medias

Para el caso de variables *continuas*, nuevamente se parte del supuesto de que la variable de interés sigue una distribución normal; por lo tanto, la estimación se basa en la diferencia de los promedios. Los factores principales que se deben considerar para calcular el tamaño de muestra son: la diferencia clínicamente importante (d), la variabilidad (S^2) y el nivel de confianza (Z_α).

En general, se utiliza la siguiente fórmula para calcular el tamaño de muestra para promedios:

$$n = 4f(\alpha,\beta)\left(\frac{s^2}{d^2}\right)$$

En el siguiente ejemplo, se quiere calcular un tamaño de muestra para estimar si la aplicación de una nueva dieta ayuda a reducir peso en personas que sufren de sobrepeso. Los investigadores esperan que la aplicación de la nueva dieta durante 6 meses reduzca en, aproximadamente, 7 kg el peso de los pacientes, comparados con un grupo control. La desviación estándar del peso de los pacientes en cada uno de los grupos se espera que sea de alrededor de 20 kg. Con un $Z\alpha$ 95 % y un poder del 90 %, el número de personas que se necesitan para el estudio es:

$$n = 4\,(8{,}56)\,\frac{20^2}{7^2} = \frac{17{,}12\times 400}{49} = 279{,}6 \sim 280\ \textit{personas}$$

Según lo planteado, en total, se necesitaría una muestra de 280 personas para hacer el estudio (140 por cada grupo).

TAMAÑO DE MUESTRA SEGÚN LOS DISEÑOS DE INVESTIGACIÓN

Como ya se mencionó, el cálculo del tamaño de muestra se halla directamente relacionado con el diseño utilizado en la investigación, el cual, a su vez, depende de los objetivos de interés, y, por ende, de la pregunta de investigación. Dada la gran cantidad de fórmulas existentes para tal fin, es necesario conocer los aspectos básicos que requiere cada una de las fórmulas para el cálculo.

Tamaño de muestra para estudios transversales

En el caso de estudios de tipo transversal, se puede usar la fórmula para una proporción. Sin

Tabla 6.1 Valores críticos más utilizados para error tipo I y poder

Poder	Error tipo I					
	Una cola			Dos colas		
	0,01	0,05	0,10	0,01	0,05	0,10
0,80	10,04	6,18	4,51	11,68	7,85	6,18
0,85	11,31	7,19	5,37	13,05	8,98	7,19
0,90	13,02	8,56	6,57	14,88	10,51	8,56

embargo, autores como Levy y Lemeshow[18], presentan una variedad de fórmulas para el caso de estimar proporciones o promedios de interés en una poblacional, pero considerando otros factores; especialmente, cuando la prevalencia de interés es baja. En el caso de estimar la prevalencia partiendo de una población de interés, se puede recurrir a la formula vista en el apartado "Tamaño de muestra para una proporción". En general, los factores que se deben considerar para el cálculo del tamaño de muestra son: la proporción del evento de interés (p), el nivel de confianza (Z_α) y la precisión buscada (ε). Los valores críticos más utilizados ya fueron presentados en la **tabla 6.1**.

La fórmula para calcular el tamaño de muestra para proporciones, cuando la prevalencia de interés es baja y la muestra se selecciona de manera aleatoria, está dada por:

$$n \geq \frac{\left(Z_{1-\frac{\alpha}{2}}\right)^2 Np(1-p)}{(N-1)\varepsilon^2 p^2 + \left(Z_{1-\frac{\alpha}{2}}\right)^2 p(1-p)}$$

Por ejemplo, se quiere estimar la proporción de personas con asma en la población colombiana. Según estudios en poblaciones anteriores, se espera que dicha proporción se encuentre en torno al 10 %. Si los investigadores quieren que tal estimación se encuentre más o menos en el 7 % en torno al estimador, con un Z_α 95 % en una población estimada de 45 millones de habitantes, el tamaño de muestra sería el siguiente:

$$n \geq \frac{(3{,}84)\,(45000000)\,0{,}10\,(0{,}90)}{(44999999)\,(0{,}0049)\,(0{,}01)+(3{,}84)\,(0{,}1)\,(0{,}9)}$$

$$= 7054{,}89 \sim 7055 \textit{ personas}$$

Por lo anterior, el tamaño de muestra mínimo que se necesitaría el estudio es de 7055 personas.

Tamaño de muestra para estudios de casos y controles

Los estudios de casos y controles se utilizan para evaluar si hay asociación entre posibles factores de exposición y un desenlace dicotómico de interés. Su medida de efecto es el *riesgo relativo indirecto* (OR, por las siglas en inglés de *odds ratio*); por lo tanto, para calcular el tamaño de muestra, la hipótesis nula de interés es que el OR es igual a 1; o sea, que no hay asociación entre la variable de exposición y el desenlace. Si se conoce la frecuencia de exposición en los casos y los controles, se puede calcular el tamaño de muestra a partir de las fórmulas estándar empleadas para comparar dos proporciones. Sin embargo, se han propuesto algunas fórmulas más generales, como las dadas por Schlesselman[19], Dupont[20], Lubin[21-23], Lui[24], Campbell[25], Machin[26] y Chow[6], por mencionar solo algunas, y en las que se consideran otros aspectos, como el número de controles por caso, etc.

Los factores principales que se deben considerar para el cálculo del tamaño de muestra son: la frecuencia de la exposición entre los controles (λ_0), la frecuencia de la exposición entre los casos ($\lambda_0+\delta$) —o, en su reemplazo, el OR esperado entre casos y controles—, la razón entre casos y controles (C), el nivel de confianza (Z_α) y el poder (Z_β).

La fórmula estándar para calcular el tamaño de muestra es:

$$n = \frac{1}{(\lambda_0-\Omega)^2}\left\{ Z_{1-\frac{\alpha}{2}}\sqrt{\left(1+\frac{1}{C}\right)\Pi(1-\Pi)} + Z_{1-\beta}\sqrt{\frac{\lambda_0(1-\lambda_0)}{C}+\Omega(1-\Omega)}\right\}^2$$

$$\Omega = \left(\frac{\lambda_0+\delta}{1+\delta}\right) y \Pi = \frac{\lambda_0}{1+C}\left[C+\frac{\Omega}{\lambda_0}\right]$$

$$m = cn \text{ y } N = m+n = n(C+1)$$

EJEMPLO

Un grupo investigador quiere evaluar si hay asociación entre el consumo habitual de bebidas alcohólicas y el hecho de sufrir un infarto de miocardio. Para ello, van a hacer un estudio de casos y controles, donde le preguntan a un grupo de pacientes que sufrieron infarto de miocardio (casos) y a un grupo de pacientes sanos (controles) si consumían a menudo bebidas alcohólicas. Si se considera que cerca del 30 % de los controles consumen a menudo bebidas alcohólicas, y se espera que los casos tengan un OR de 2 respecto a los controles, el tamaño de muestra mínimo necesario, calculado con un nivel de confianza del 95 %, un poder del 80 % y una razón entre casos y controles de 1, sería de:

$$\Omega = \left(\frac{\lambda_0 - \delta}{(1+\delta)}\right) = \frac{0{,}3+0{,}3}{1+0{,}3} = \frac{0{,}6}{1{,}3} = 0{,}4615$$

$$\Pi = \frac{0{,}3}{(1+1)}\left[1+\frac{0{,}4615}{0{,}3}\right] = 0{,}15(2{,}5385) = 0{,}3808$$

$$n = \frac{1}{(0{,}3-0{,}4615)^2}\left\{1{,}96\sqrt{(2)0{,}3898(0{,}6192)} + 0{,}84\sqrt{0{,}3\,(0{,}7)+0{,}4615\,(0{,}5385)}\right\}^2$$

$$n = 38{,}34\,(1{,}9148^2) = 140{,}56 \sim 141 \textit{ casos}$$

Según lo descrito, en total se necesitaría una muestra de 282 pacientes: 141 casos y 141 controles.

Tamaño de muestra para estudios de cohortes

Los estudios de cohortes se utilizan para evaluar si hay asociación causal entre posibles factores de riesgo y un desenlace dicotómico de interés. Su medida de efecto es el ***riesgo relativo*** **(RR); por lo tanto, para calcular el tamaño** de muestra, la hipótesis nula de interés es que el **RR** es igual a 1; o sea, que no hay asociación entre los factores de riesgo y el desenlace de interés. Al igual que para los estudios de casos y controles, si se conoce la frecuencia de exposición, se puede calcular el tamaño de muestra a partir de las fórmulas estándar empleadas para comparar dos proporciones. No obstante, se han propuesto fórmulas más específicas para dicho cálculo, como las presentadas por Fleiss[27], Lachin[28], Altman[29], Tubert-Bitter[30], Dupont[31], Jianwen[32], Whitley[33] y Doll[34], por mencionar solo algunos, y en las que se consideran otras opciones, como el número de controles por caso, entre otras.

Los factores principales que se deben considerar para el cálculo del tamaño de muestra son: la frecuencia de la exposición entre los controles (λ_0), la frecuencia de la exposición entre los casos ($\lambda_0+\delta$) —o en su reemplazo, el **RR esperado entre casos y controles—, la razón** entre casos y controles (C), el nivel de confianza (Z_α) **y el poder** (Z_β).

La fórmula estándar para calcular el tamaño de muestra es:

$$n = \frac{\left\{Z_{1-\frac{\alpha}{2}}\sqrt{(C+1)\bar{\lambda}(1-\bar{\lambda})} + Z_{1-\beta}\sqrt{\lambda_0(1-\lambda_0)+C(\lambda_0+\delta)(1-\lambda_0-\delta)}\right\}^2}{C(\delta^2)}$$

$$\bar{\lambda} = \frac{[C\lambda_0 + (\lambda_0+\delta)]}{1+C}$$

$$m = cn \text{ y } N = m+n = n(C+1)$$

Retomando el ejemplo anterior, se planea un estudio de cohorte donde se seleccionará como controles a pacientes que no consumen a menudo bebidas alcohólicas, y como casos, a quienes sí las consumen; se siguen en el tiempo ambos grupos hasta evaluar si sufren o no un infarto miocárdico. Si se supone que el 20 % de los casos y el 12 % de los controles sufrirán infarto de miocardio, ¿cuál debería ser el tamaño de muestra mínimo necesario, si se quieren un nivel de confianza del 95 %, un poder del 80 % y una razón entre casos y controles de 1?

$$\bar{\lambda} = \frac{[0{,}12+0{,}20]}{2} = 0{,}16$$

$$n = \frac{\left\{1{,}96\sqrt{(2)\,0{,}16\,(0{,}84)} + (0{,}84)\sqrt{0{,}12\,(0{,}88)+(0{,}20)\,(0{,}80)}\right\}^2}{(0{,}08^2)}$$

$$n = \frac{2{,}0998}{0{,}0064} = 328{,}10 \sim 329 \textit{ casos}$$

Según lo anterior, en total se necesitaría una muestra de 658 pacientes: 329 casos y 329 controles.

TAMAÑO DE MUESTRA PARA PRUEBAS DIAGNÓSTICAS

En esta clase de estudios se busca determinar la calidad de una prueba diagnóstica que se usa para evaluar la presencia o no de cierta condición o característica en un paciente. Indicadores como la sensibilidad, la especificidad y los valores predictivos positivos y negativos son las medidas tradicionales que ayudan a evaluar la validez y la seguridad de dichas pruebas.

Es importante recalcar que, además de los indicadores anteriores, un factor importante en el cálculo del tamaño de muestra es la ***prevalencia de*** *la característica de interés en la población de estudio*. Por ello, en algunos casos se recurre a otros índices que son bastante útiles y no dependen de la prevalencia, y que corresponden a las ***razones de probabilidad***, las cuales reúnen en un solo índice la sensibilidad y la especificidad de la prueba.

El cálculo del tamaño de muestra depende, específicamente, del objetivo de interés de la investigación, por lo que existen fórmulas específicas para evaluar la sensibilidad, la especificidad y las razones de verosimilitud de una prueba, así como para establecer ***áreas bajo la curva*** (ABC), o la comparación de curvas de características operativas del receptor (curvas COR)[26,35-37].

Si se quiere evaluar una prueba diagnóstica a través de sensibilidad y especificidad, la fórmula utilizada para calcular el tamaño de muestra es la misma que se presentó en el apartado "Tamaño de muestra para una proporción", de una proporción y el tamaño de muestra que se obtiene es el necesario por grupo.

Para el caso de la comparación de dos pruebas diagnósticas en grupos independientes a través de sensibilidad o especificidad, la fórmula utilizada para calcular el tamaño de muestra es la misma que se presentó en el apartado "Tamaño de muestra para diferencia de proporciones", de diferencia de dos proporciones independientes; por tanto, los parámetros que se consideran para su cálculo son: el error tipo I, el error tipo II, la sensibilidad y la especificidad y la prevalencia esperada de la característica de interés.

Tamaño de muestra para modelos de regresión

Debe recordarse que la indicación más común para estimar un tamaño de muestra es la ***evaluación de una hipótesis*** o la ***estimación*** *de un* IC, por lo cual el enfoque se centra en la exactitud en torno al estimador. La búsqueda de poder en el estudio, o de precisión, debe considerarse también en el cálculo de tamaño de muestra para un modelo de regresión, que también se plantea hipótesis y busca que sus estimadores sean válidos y precisos[38-41].

Las consideraciones para calcular un tamaño de muestra en modelos de regresión múltiple deben incluir el poder estadístico para lograr el ajuste global del modelo, así como para variables explicativas específicas, y, en algunos casos, según el tipo de modelo (asociación o predicción) y el tipo de variable de respuesta (dicotómica, ordinal, multinomial, continua), será necesario obtener cierta precisión y exactitud en alguno de los parámetros del modelo.

No es de extrañar, dadas las diferencias entre los posibles modelos y el tipo de desenlace evaluado, que los tamaños de muestra obtenidos sean diferentes entre sí. Adicionalmente, este puede variar entre modelos de regresión simple o múltiple. En ese contexto, un tamaño de muestra óptimo para un tipo de modelo puede no ser suficiente para otro tipo de modelo.

Tamaño de muestra para modelos de regresión lineal simple con variable de exposición dicotómica

Para un modelo de regresión lineal simple, donde la variable de desenlace es continua y la variable independiente es dicotómica, y asumiendo que se cumplen los supuestos de normalidad de la variable de respuesta y el de igualdad de varianzas poblaciones en cada uno de los grupos, se puede usar el cálculo de tamaño de muestra para dos promedios independientes (apartado "Tamaño de muestra para diferencia de medias").

Tamaño de muestra para modelos de regresión lineal simple con variable de exposición continua

Cuando el modelo de regresión también es lineal, pero la variable independiente es continua y se cumple el supuesto de normalidad de la variable de respuesta, la hipótesis de que la

pendiente es igual a cero equivale a probar que no hay correlación lineal entre estas variables, y que, por lo tanto, el coeficiente de correlación es cero ($\rho = 0$). En tal caso, el tamaño de muestra mínimo necesario se puede obtener a través de la siguiente fórmula[28,42-45]:

$$n = \left[\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{C(p)}\right]^2 + 3$$

Donde, $C(p) = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$ conocida como la transformación Z de Fisher[46,47].

Como ejemplo, supongamos que se quiere estimar el tamaño de muestra para evaluar si hay asociación entre el nivel de colesterol LDL —por las siglas en inglés de *low density lipoprotein*— (variable respuesta) y los triglicéridos (X) en una población de sujetos diabéticos. Con un nivel de confianza del 95 %, un poder del 80 %, y una correlación estimada de 0,20 se obtiene:

$$n = \left[\frac{1{,}96 + 0{,}8416}{0{,}20}\right]^2 + 3 = 196{,}22 \sim 197 \text{ pacientes}$$

Por lo evidenciado, el tamaño de muestra mínimo necesario es de 197 pacientes, con información del nivel de colesterol LDL y triglicéridos.

Tamaño de muestra para modelos de regresión lineal múltiple con variable de exposición continua

En términos más generales, en el caso de modelos de regresión lineal múltiple, donde hay una sola variable independiente de exposición, de tipo continuo, y las demás variables independientes son variables de control, el tamaño de muestra puede obtenerse mediante una aproximación de la formula anterior[40,41]. El mínimo necesario de observaciones para tal situación es el siguiente:

$$n_m = \frac{n_s}{1 - \rho^2_{x_1|(x_2,\ldots,x_p)}}$$

Donde n_s es el tamaño de muestra para una regresión lineal simple y ρ^2 es el coeficiente de determinación entre la variable de exposición y las variables de control.

Retomando el ejemplo anterior, si se quiere estimar un tamaño de muestra para evaluar si hay asociación en una población de pacientes diabéticos entre el nivel de colesterol LDL (variable respuesta) y los triglicéridos (X_1), controlando por edad, sexo e índice de masa corporal (IMC), con una confiabilidad del 95 %, un poder del 80 %, una correlación estimada entre la variable respuesta y la variable exposición de 0,20, y un coeficiente de determinación entre la variable exposición y las variables de control de 0,30, el tamaño de muestra mínimo necesario es de:

$$n_m = \frac{197}{1 - 0{,}30} = 281{,}4 \sim 282 \text{ pacientes}$$

Otra fórmula utilizada para determinar el tamaño de muestra mínimo necesario en un modelo de regresión lineal múltiple[17,38,48], donde se tiene una principal variable de exposición, es la siguiente:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}}{d}\right)^2 \left(\frac{1 - R^2}{1 - R^2_{xyj}}\right) + p + 1$$

En este caso, el tamaño de muestra se basa en hallar una precisión buscada (d) del IC obtenido para el coeficiente de regresión estandarizado de la variable de exposición, controlando por las demás covariables. Así, Z corresponde al valor crítico de una distribución normal estándar, d corresponde a la precisión buscada, p corresponde al número de predictores (variables independientes o covariables) del modelo, R^2 es el coeficiente de determinación entre la variable respuesta Y y las variables independientes, y R^2_{XXj} representa el coeficiente de correlación múltiple entre la variable exposición y las demás variables independientes

Nuevamente utilizando el ejemplo anterior, se quiere estimar un tamaño de muestra para evaluar si hay asociación en una población de sujetos diabéticos entre el nivel de colesterol LDL (variable respuesta) y los triglicéridos (X_1), controlando por edad, sexo e IMC. Con una confiabilidad del 95 %, un poder del 80 %, una precisión de 0,10, un coeficiente de determinación entre la variable respuesta y la variable exposición de 0,20, y un coeficiente de determinación entre la variable exposición y las variables de control de 0,30, el tamaño de muestra mínimo necesario es:

$$n = \left(\frac{1{,}96}{0{,}10}\right)^2 \left(\frac{1 - 0{,}20}{1 - 0{,}30}\right) + 4 + 1 = 444{,}1 \sim 445 \text{ pacientes}$$

Tamaño de muestra para modelos de regresión logística simple con variable de exposición dicotómico

En el caso de un modelo de regresión logística simple, cuya variable de respuesta y cuya variable de exposición son dicotómicas, se puede usar la fórmula para la comparación de dos proporciones independientes (apartado "Tamaño de muestra para diferencia de proporciones").

Como ejemplo, supóngase que se quiere estimar un tamaño de muestra para evaluar si hay asociación entre la presencia de diabetes ***mellitus*** tipo 2 (variable respuesta) y la obesidad (X) en una población de pacientes diabéticos. Se espera que la proporción de personas que presente diabetes ***mellitus*** **tipo 2 en los sujetos no obesos** esté alrededor del 7 %, y que la proporción de sujetos obesos con diabetes ***mellitus*** tipo 2 sea del 14 %. Con un Z_α 95 % y un poder del 80 %, el tamaño de muestra sería el siguiente:

$$n = 2\left[\frac{0{,}14\,(0{,}86) + 0{,}07\,(0{,}93)}{(0{,}14 - 0{,}07)^2}\right](6{,}18) = 467{,}91 \sim 468 \text{ pacientes}$$

Según lo visto, el tamaño de muestra mínimo total que se necesitaría en el estudio es de 468 pacientes: 234 obesos y 234 no obesos.

Tamaño de muestra para modelos de regresión logística simple con variable de exposición continua

En caso de un modelo de regresión logística simple, pero donde la variable independiente de exposición es continua, el tamaño de muestra mínimo necesario puede obtenerse a partir de la fórmula utilizada para comparar dos proporciones independientes (apartado "Tamaño de muestra para diferencia de proporciones"), pero haciendo un ajuste a través de la fórmula dada en el apartado "Tamaño de muestra para modelos de regresión lineal simple con variable de exposición continua"[41].

Volvamos el ejemplo anterior, para evaluar si existe asociación entre la presencia de diabetes ***mellitus*** **tipo 2 (variable respuesta) y la** obesidad (X), controlando por algunas covariables como edad y sexo. Con una proporción de personas que presentan diabetes ***mellitus*** tipo 2 en los sujetos no obesos del 7 %, una proporción de sujetos obesos con diabetes *mellitus* **tipo 2 del 14 %, un Z_α 95 %, un poder** del 80 % y un coeficiente de determinación entre la variable exposición y las variables de control de 0,15, el tamaño de muestra sería el siguiente:

$$n_m = \frac{468}{1 - 0{,}15} = 550{,}59 \sim 552 \text{ pacientes}$$

Por lo anterior, el tamaño de muestra mínimo total que se necesitaría en el estudio es de 552 pacientes: 276 obesos y 276 no obesos.

Tamaño de muestra para modelos de regresión logística múltiple

Existen en la actualidad diferentes enfoques para calcular el tamaño de muestra en modelos de regresión logística múltiple. Algunos se basan en el número de parámetros y en la precisión en torno a un coeficiente[49-51], y otras, en el poder estadístico[40-51-53]. En el primer caso, una de las recomendaciones que muchos autores dan es contar con al menos 10 desenlaces o casos por parámetro en el modelo, si bien algunos autores sugieren al menos 20[50,54].

Lo anterior significa que en caso de contar con k parámetros en el modelo, se necesitarían (cpp) **× (k + 1) casos, lo cual debe ajustarse por** la probabilidad de obtener un desenlace o un caso. Por tanto, la formula está dada por:

$$n = \frac{(k+1) \times (cpp)}{P_{Menor}}$$

Por ejemplo, si la prevalencia de un evento de interés es del 10 %, con 12 parámetros en el modelo y esperando obtener 10 casos por parámetro, el tamaño de muestra mínimo necesario es:

$$n = \frac{(13) \times (10)}{0{,}1} = 1300 \text{ observaciones}$$

CONSIDERACIONES FINALES

Sin duda, el tema del cálculo del tamaño de muestra tiende a ser complejo y mal entendido, y, más preocupante aún, muchos estudios no consideran dicho aspecto, o, simplemente, no hacen un cálculo formal de él.

Hacer un cálculo apropiado es imprescindible para cualquier diseño de investigación, pero todavía más importante es la transparencia en el momento de calcularlos, de tal manera que se justifique adecuadamente el porqué de los parámetros utilizados.

Aunque algunas de las fórmulas (y, por ende, algunos de los cálculos) parecen ser sencillas y simples, no siempre son comprensibles para los investigadores; por ello, se recomienda contar con el apoyo de un grupo interdisciplinario, donde se tenga la oportunidad de disponer de un estadístico o de una persona capacitada para tal función. El uso de tablas, en muchos casos, puede facilitar el cálculo.

Aunque algunos cálculos pueden conducir a tamaños de la muestra en apariencia muy pequeños, ello no significa que no se puedan hallar resultados estadísticamente significativos; y, viceversa, los tamaños de muestra grandes no aseguran diferencias clínicamente significativas.

Es importante evaluar cuán precisa es la estimación obtenida, para tener tranquilidad sobre los resultados que se encuentren. Si se descuida dicho aspecto, se puede estar cometiendo un grave error, apoyado en resultados nada confiables.

La planeación del cálculo del tamaño de muestra es un paso importante, y, a la vez, exigente, que se debe considerar en una fase temprana de cualquier proyecto de investigación. Por lo tanto, es imprescindible tener información buena y confiable, además de saber cómo emplearla adecuadamente en el cálculo.

Referencias

1. Rondón MA, Rodríguez VA. Algunos conceptos básicos para el cálculo del tamaño de muestra. Univ Med. 2007;48:334-9.
2. Dennis RJ, Pérez A. Cálculo del tamaño de muestra. En: Morales AR, Zárate LEM, editores. Epidemiología clínica: Investigación clínica aplicada. Bogotá: Médica Panamericana; 2004. p. 141-62.
3. Dennis RJ. Cómo estimar el tamaño de muestra en investigaciones con humanos. Acta Medica Colomb. 1989;14:92-9.
4. Columb MO, Stevens A. Power analysis and sample size calculations. Curr Anaest Crit Care. 2008;12–4.
5. Daly LE. Confidence intervals and sample sizes: don't throw out all your old sample size tables. BMJ Br Med J. 1991;302:333-6.
6. Chow S-C. Sample size calculations for clinical trials. Wiley Interdiscip Rev Comput Stat. 2011;3:414-27.
7. Hulley SB, Cummings SR, Browner WS. Diseño de investigaciones clínicas. Philadelphia: Lippincott Williams & Wilkins; 2007.
8. Lwanga S, Lemeshow S. Sample size determination in health studies: A practical manual, 1991. Ginebra: World Health Organization; 1991.
9. Lenth RV. Some practical guidelines for effective sample size determination. Am Stat. 2008;55:187-93.
10. Dawson-Saunders B, Trapp RG, de la Vega Sánchez ML. Bioestadística médica México: El Manual Moderno; 2005.
11. Daniel W. Biostatistics. library.uru.ac.th [internet]. 1999 [citado: 2015 ene. 20]. Disponible en: http://library.uru.ac.th/stat/images/cont12075.pdf\nfile:///Users/sanghyuk/Documents/Library.papers3/Articles/1999/Daniel/library.uru.ac.th 1999 Daniel.pdf\npapers3://publication/uuid/757B1016-E855-477D-A561-1ECEB16BA3F2
12. Milton JS. Estadística para biología y ciencias de la salud. Madrid: McGraw-Hill Interamericana de España; 2007.
13. Chuang-Stein C, Kirby S, Hirsch I, et al. The role of the minimum clinically important difference and its impact on designing a trial. Pharm Stat. England; 2011;10:250-6.
14. Cohen JE. The statistical power of abnormal-social psychological research: a review. J Abnorm Soc Psychol. 1962;65:145-53.
15. Norman GR, De Burgh Norman JE, et al. Bioestadística [internet]. Harcourt; 1996. [citado 2015 ene. 20]. Disponible en: http://books.google.com.co/books?id=yCLUYL14tlQC
16. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ. 2001;322:226-31.
17. Maxwell SE, Kelley K, Rausch JR. Sample size planning for statistical power and accuracy in parameter estimation. Annu Rev Psychol. 2008;59:537-63.
18. Levy PS, Lemeshow S. Sampling of populations: Methods and applications. Technometrics. 1999:1-575 p.
19. Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. Am J Epidemiol. 1974;99:381-4.
20. Dupont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. Control Clin Trials. 1998;19:589-601.
21. Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. Am J Epidemiol. 1990;131:552-66.

22. Lubin JH. Power and sample size calculations in case-control studies of gene- environment interactions: Comments on different approaches. Am J Med Genet A. 2011;155:C1.
23. Lubin JH, Gail MH, Ershow AG. Sample size and power for case-control studies when exposures are continuous. Stat Med. 1988;7:363-76.
24. Lui KJ, Cumberland WG. Sample size determination for equivalence test using rate ratio of sensitivity and specificity in paired sample data. Control Clin Trials. 2001;22:373-89.
25. Campbell MK, Thomson S, Ramsay CR, et al. Sample Size Calculator. Comput Biol Med. 2004;113-25.
26. Machin D, Campbell M, Fayers, PPA. Sample size tables for clinical studies. 2nd ed. Oxford: Blackwell; 1997.
27. Fleiss JL, Levin B, Paik MC, et al. Comparative studies: prospective and retrospective sampling. Statistical methods for rates and proportions [internet]. 2004 [citado 2015 ene. 20]. Disponible en: http://dx.doi.org/10.1002/0471445428.ch7
28. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Cont Clin Trials. 1981;93-113.
29. Altman DG. Statistics and ethics in medical research: III How large a sample? Br Med J. 1980;281:1336-8.
30. Tubert-Bitter P, Bégaud B, Moride Y, et al. Sample size calculations for single group post-marketing cohort studies. J Clin Epidemiol. 1994;47:435-9.
31. Dupont WD. Power calculations for matched case-control studies. Biometrics. 1988;44:1157-68.
32. Cai J, Zeng D. Sample size/power calculation for case-cohort studies. Biometrics. 2004;1015-24.
33. Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care. 2002;6:335-41.
34. Doll R. Cohort studies: history of the method. I. Prospective cohort studies. Soz Praventiv Med. 2001;46:75-86.
35. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol. 1991;44:763-70.
36. Carley S, Dosman S, Jones SR, et al. Simple nomograms to calculate sample size in diagnostic studies. Emerg Med J. 2005;22:180-1.
37. Obuchowski NA. Sample size calculations in studies of test accuracy. Stat Methods Med Res. 1998;7:371-92.
38. Kelley K, Maxwell SE. Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. Psychol Methods. United States; 2003 Sep;8(3):305–21.
39. Demidenko E. Sample size determination for logistic regression revisited. Stat Med. 2007;26:3385-97.
40. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Stat Med. 1998;17:1623-34.
41. Kleinbaum D, Kupper L, Nizam A, et al. Applied regression analysis and other multivariable methods [Internet]. Cengage Learning; 2013 [citado 2015 ene. 20]. Disponible en: https://books.google.com.co/books?id=v590AgAAQBAJ
42. Pan W. Sample size and power calculations with correlated binary data. Control Clin Trials. 2001;22:211-27.
43. Rothstein HR, Borenstein M, Cohen J, et al. Statistical power analysis for multiple regression/correlation: A computer program. Educ Psychol Measur. 1990;819-30.
44. Gatsonis C, Sampson AR. Multiple correlation: exact power and sample size calculations. Psychol Bull. 1989;106:516-24.
45. Schönbrodt FD, Perugini M. At what sample size do correlations stabilize? J Res Pers. 2013;47:609-12.
46. Fisher RA, Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population [internet]. Biometrika. 1915. p. 507-21 [citado 2015 ene. 20]. Disponible en: http://biomet.oxfordjournals.org/cgi/reprint/10/4/507.pdf
47. Ozer DJ. Correlation and the coefficient of determination. Psychol Bull. 1985 ;307-15.
48. Maxwell SE. Sample size and multiple regression analysis. Psychol Methods. 2000;5:434-58.
49. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. J Clin Epidemiol. 1995;48:1495-501.
50. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373-9.
51. Whittemore AS. Sample size for logistic regression with small response probability. J Am Stat Assoc [internet]. 1981;76:27 [citado 2015 ene. 20]. Disponible en: http://www.jstor.org.gate.lib.buffalo.edu/stable/info/2287036
52. Hsieh FY. Sample size tables for logistic regression. Stat Med. 1989;8:795-802.
53. Cohen J. Statistical power analysis. Current directions in psychological science. 1992;98-101.
54. Hosmer DW, Lemeshow S. Applied logistic regression [internet]. Wiley series in probability and statistics Texts and references section. 2000. 392 p. [citado 2015 ene. 20]. Disponible en: http://www.amazon.com/dp/0471356328

CAPÍTULO 7

Variables no observables en epidemiología clínica: el uso de constructos

Socorro Moreno Luna

INTRODUCCIÓN

Cada vez más a menudo, los clínicos están interesados en determinar la satisfacción con la salud o con la atención recibida, la adherencia a las recomendaciones formuladas o la calidad de vida de sus pacientes. En muchas intervenciones terapéuticas se intenta modular o modificar las creencias, las actitudes y las preferencias de los sujetos frente a múltiples factores del continuo salud-enfermedad, pero ¿a qué corresponden estos conceptos?, ¿cómo evaluarlos o medirlos si no se puede acceder a ellos de manera directa?, ¿las definiciones de diccionarios o las provistas por organismos internaciones son suficientes? Estas preguntas se relacionan con conceptos abstractos, de los cuales tenemos alguna evidencia empírica, pero esta, a menudo, es incompleta o confusa.

¿QUÉ ES UN CONSTRUCTO?

El concepto de constructo se ha ligado tradicionalmente al desarrollo de sistemas o de instrumentos de medición. Sin embargo, la comprensión de esos conceptos en los que estamos interesados al hacer investigación incluye, pero también excede, el ámbito de la medición.

Puede definirse un constructo como una variable explicativa que no puede observarse directamente. En una definición más completa, un constructo es una abstracción teórica que permite interpretar una serie de comportamientos, tales como la manifestación de rasgos, de atributos o de procesos, que ocurren en una red de asociaciones y pueden ser generalizables más allá de las características de personalidad o de temperamento de los sujetos que están siendo observados[1,2].

Ejemplo: Se busca relacionar la inteligencia con la capacidad para tener éxito profesional. Sería fácil decir que cuanto más inteligente sea una persona mayor será su probabilidad de éxito. La medición de la inteligencia, sin embargo, enfrenta la formidable dificultad de que esta característica no puede observarse de forma directa. Y como se trata de una abstracción mental, deben proponerse componentes (rasgos, signos, atributos) que la puedan representar.

Los constructos pueden referirse a ideas (autoestima, tabú, eutanasia), a aspectos sociales (racismo, filantropía, etnicidad), a aspectos de las organizaciones (responsabilidad social, cultura organizacional, alianzas) o, incluso, a cosas más concretas (células madre). La clasificación por sexo es un buen ejemplo de constructo, pues no se trata de la simple separación de

las personas en hombres o mujeres. Hay otros conceptos complejos involucrados, o que deben tenerse en cuenta, como homosexual, bisexual, transexual, etc.

Piense por un momento qué quiere hacer: ¿controlar o modificar un comportamiento? Cuando se plantean intervenciones para modificar la adherencia a los tratamientos, ¿es una modificación, o se busca la asociación de una serie de comportamientos? Cuando se plantea una relación entre la personalidad vigorosa y la susceptibilidad al estrés laboral, ¿puede decirse que se busca describir a un grupo de personas según sus comportamientos? Cuando se quiere caracterizar las actitudes de un grupo de personas frente a la apariencia física de los médicos, se buscan las partes observables (***comportamientos***) de los cuales un sujeto puede dar cuenta o que un observador puede registrar. Lo importante es estar seguro de que esos comportamientos reflejan realmente en lo que usted está interesado.

Un constructo debe estar tan bien definido como para que no haya espacio para interpretaciones erróneas, y debe quedar tan definido que sea reproducible o utilizable en otros estudios; además, cabe aclarar que un mismo constructo puede, a pesar de todo, tener interpretaciones diferentes o explicaciones diversas. Dichos conceptos son fundamentales para la validación de escalas que utilicen constructos.

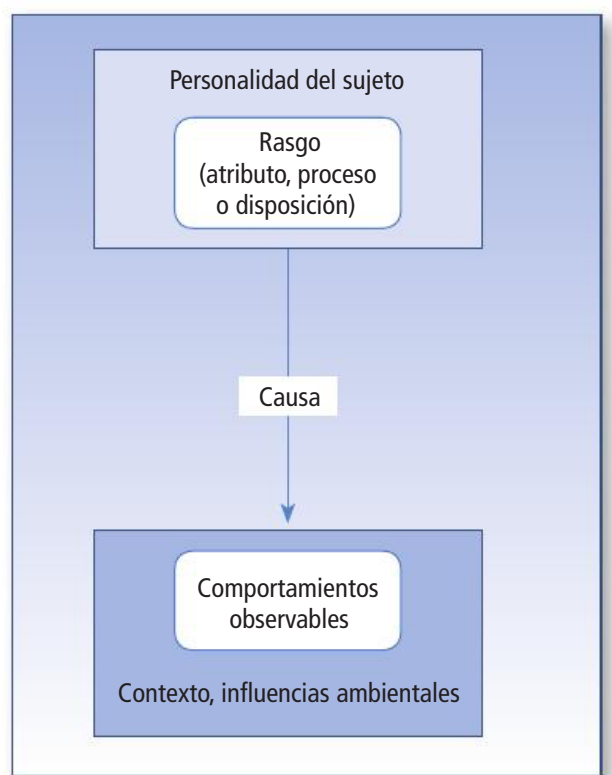
La definición precisa de un constructo implica que sus conceptos hayan sido expresados (o hechos explícitos), y así puedan ser criticados, relacionados con otros constructos, definidos operacionalmente y probados (lo cual implica que sean ***mensurables***)[1].

Un constructo no es una variable: se usan variables para operacionalizarlo (a veces, una sola, como en el caso del sexo). Puede decirse que los constructos deben ser traducidos de lo abstracto a lo concreto, de formas mensurables (***variables***). Piense el lector en el constructo *felicidad*, que puede ser asociado a amor, a tranquilidad, a seguridad financiera, a ciertos alimentos, a flores, a música o a posesiones (dinero, joyas, vehículos). La traducción a lo concreto no es fácil, y necesita claridad y explicación.

Ahora bien, ¿cómo se puede estar seguro de que lo que se observa realmente se corresponde con un concepto intangible? Es indispensable ubicar el concepto (digamos, ***adherencia***) dentro de un contexto particular; por ejemplo, tratamientos para entidades crónicas. Y a partir de una perspectiva del mundo, un paradigma, como el del empoderamiento del paciente. Con ello en mente, uno podría plantear el **esquema 7.1**.

La personalidad del sujeto embebe una serie de rasgos, que pueden ser atributos, procesos o disposiciones (en el caso de la adherencia, se trata de procesos de autocuidado) que causan comportamientos observables (como usar una alarma para tomar a tiempo el medicamento) que ocurren en un contexto particular, son objeto de influencias ambientales y pueden sufrir modificaciones según si el sujeto está siendo observado intencionalmente, o si se están manipulando sus condiciones (si el paciente forma parte de un grupo de soporte para su enfermedad, por ejemplo). La perspectiva de un paciente empoderado que decide cómo y cuáles actividades de salud realiza para poner en marcha su propio plan de salud modula la identificación, el registro y la evaluación de los comportamientos del sujeto y de su interacción con el sistema de salud.

Como el lector puede suponer, existe una amplia serie de comportamientos (ya sean direc-



Esquema 7.1 Relación entre rasgos y comportamientos observados.
Fuente: Elaboración propia.

[1] Laerd dissertation. http://dissertation.laerd.com/getting-started.php (consultado el 4 de abril de 2015).

tamente observables o que el sujeto reporta) que podrían relacionarse con un mismo constructo. Teniendo eso en mente, surge la inquietud de cómo saber (o garantizar) que la selección de algunos comportamientos representa realmente el constructo de interés. Para resolverla, se propone un proceso gradual de acumulación de información, tanto teórica como empírica, que permita asegurar, con relativa confianza, que se está abordando el constructo de interés.

PASOS PARA DEFINIR UN CONSTRUCTO

El primer paso es hacer una búsqueda exhaustiva de la literatura teórica sobre el tema, para reunir los diferentes paradigmas sobre el concepto. En el caso de nuestro ejemplo, significa buscar las aproximaciones que los teóricos del comportamiento (psicólogos, enfermeras, antropólogos, sociólogos, etc.) han hecho sobre el concepto de adherencia al tratamiento médico. El sitio ideal para llevar a cabo dicha búsqueda está un poco más allá de las bases de literatura médica tradicionales, y abarca bases de literatura en ciencias humanas y del comportamiento (JSTOR, Education Research Information Center [ERIC], Dissertations and Theses, A&I o Psychology and Behavioral Sciences Collection).

En ocasiones, se hace tentador omitir este paso, pues el investigador puede considerar que tiene una buena impresión subjetiva sobre lo que es el constructo[1]. No se recomienda tal omisión, que puede conducir a callejones sin salida, pues en ausencia de un marco adecuado no puede haber certeza sobre cuáles de los comportamientos propuestos pueden formar parte del constructo.

Una vez se ha hecho esta búsqueda, se deben considerar en detalle y discutirse las descripciones que se han encontrado, para así asumir la perspectiva teórica más acorde con la impresión subjetiva del investigador, y que será la que guiará el proceso de definición del constructo.

Como los constructos no ocurren en el vacío, es indispensable plantear la red de relaciones entre el constructo de interés y otros constructos; o sea, describir la ***red nomológica de asociaciones*** entre comportamientos propios del constructo y comportamientos causados por otros constructos[1]. Durante dicho paso, el investigador reunirá información tanto de su propio constructo de interés como de los constructos con los cuales las teorías han mostrado alguna relación. En el ejemplo que se mencionó, adherencia se relaciona con alfabetización en salud, así que durante la búsqueda de literatura se definirán ambos constructos y se separarán los comportamientos propios (en algunos casos, exclusivos) de cada uno de ellos, y se identificará en cuáles hay una superposición. Con esto se podrá graficar una serie de relaciones entre dichos comportamientos (como un mapa conceptual) y se empezará a reconocer cuáles de ellos son fácilmente observables en las condiciones en las que se tendrá contacto con los sujetos.

De esta manera, se habrá avanzado lo suficiente como para pasar al segundo paso. Tal paso implica, como lo sugiere De Vellis, determinar qué tan específica y delimitada será la definición del constructo[1]. No es lo mismo considerar la adherencia a los tratamientos para entidades agudas que a entidades crónicas, como tampoco la adherencia en niños, en adultos o en ancianos, o en personas sin ningún nivel de escolaridad. Como puede verse, la división en la definición del constructo se basa en la utilidad tanto teórica como empírica de esta segmentación.

El siguiente paso es reunir información con los sujetos sobre su experiencia con el fenómeno. Ello significa obtener información empírica sobre los comportamientos que han sido causados por el rasgo y contrastar los hallazgos teóricos obtenidos de la revisión de literatura, con las experiencias observables que son moduladas por un contexto que influye de forma particular sobre los sujetos.

Los métodos ideales para recolectar esta información son los propios de los métodos cualitativos. Tanto la entrevista como los grupos focales pueden ser de gran utilidad al momento de reunir dicha información. Es importante que el investigador tenga para ese momento un mapa conceptual claro sobre el constructo, de tal forma que enfoque sus esfuerzos en indagar las expresiones del constructo que son más susceptibles de influencia ambiental. Esta fase corresponde a la ***operacionalización en contexto***[3]; o sea, a caracterizar los comportamientos observables causados por el rasgo que se está evaluando.

Al llevar a cabo este paso, es importante considerar la aproximación teórica que tendrá el proceso de indagación, mediada, a su vez,

por la teoría dentro de la cual se enmarca el constructo y por la aproximación metodológica que empleará el investigador.

De forma amplia, el investigador puede aproximarse al constructo como un fenómeno del cual un sujeto puede dar cuenta; o sea, evocar la experiencia individual para extraer de allí los comportamientos más representativos del constructo y el significado que los sujetos les asignan a tales comportamientos. Esta aproximación es de tipo fenomenológico.

Una vez se ha traducido el constructo en experiencias observables, se hace necesario comprender los factores que modulan el proceso, así como las relaciones que se establecen entre ellos; o sea, además de los comportamientos y de sus significados, establecer cómo se relacionan en términos de causas, consecuencias, contextos y contingencias, para articular una teoría en torno al constructo de interés. Dicha aproximación es del tipo de la teoría fundamentada. Para los lectores interesados en estas aproximaciones, se recomienda el texto introductorio de Green & Thorogood[4], en pro de entender el aporte de la investigación cualitativa en salud, y el de Dew[5], que presenta las guías a las metodologías cualitativas para el investigador clínico.

Al final del proceso descrito, el investigador dispondrá de una red de constructos (*red nomológica*) que le permitirá determinar con claridad su variable de interés, más allá de la simple denominación. Podrá saber qué quiere explorar, qué comportamientos debe registrar, qué variables debe controlar y cómo se relacionan con su variable de interés, y cuáles comportamientos puede excluir de su modelo sin que ello afecte el proceso de responder su pregunta.

La conclusión que quisiera que el lector tuviese en este punto es que, al considerar conceptos subjetivos a la hora de hacer preguntas clínicas, se debe sobrepasar la creencia de que se tiene claro qué se quiere explorar, para poder conceptualizar claramente las variables fundamentales de la pregunta de investigación.

Referencias

1. DeVellis RF. Scale development: Theory and applications. 3rd ed. Thousand Oaks: Sage Publications; 2011.
2. Messick S. Constructs and their vicissitudes in educational and psychological measurement. Psychol Bull. 1981;89:575-88.
3. Vogt DS, King DW, King LA. Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. Psychol Assess. 2004;16:231-43.
4. Green J, Thorogood N. Qualitative methods for health research. 3rd ed. Thousand Oaks: Sage Publications; 2013.
5. Dew K. A health researcher's guide to qualitative methodologies. Aust N Z J Public Health. 2007;31:433-7.

CAPÍTULO 8

Adaptación cultural y validación de escalas

Carlos Gómez-Restrepo • Martín Alonso Rondón Sepúlveda
María Beatriz Ospina García

INTRODUCCIÓN

Es habitual que en investigación se deban medir diferentes condiciones (síndromes, constructos) o variables, muchas de las cuales son desenlaces relevantes. Algunas de estas medidas son muy sencillas, ya que hay mecanismos directos y objetivos para llevarlas a cabo, como en el caso de la medición de estatura con una cinta métrica, la evaluación de la anemia con un cuadro hemático, o la del peso mediante una báscula en kilogramos. Otras variables o condiciones, tales como la calidad de vida, la depresión, la ansiedad, el tipo de personalidad, el dolor o la satisfacción con la vida, son más abstractas, más complejas, y no se pueden medir directamente. Muchas de ellas se componen de un mayor número de categorías, dominios o factores que las representan, por lo cual no tienen una medida directa y requieren mediciones especiales (escalas) que contemplen su complejidad.

Los proyectos de investigación que deben evaluar las mencionadas condiciones o variables recurren a menudo a escalas o instrumentos de medición para evaluarlas[1]. En este capítulo se profundizará en la adaptación y la validación de escalas existentes, así como en el desarrollo *de novo* de algunas de ellas en caso de que no existan o de que las existentes no sean adecuadas para el propósito del estudio. Para este último caso, se debe hacer un proceso que culmina con la selección de una serie de preguntas relevantes (ítems) tras un proceso de adaptación y validación.

Una ***escala*** **(instrumento de medición) se** define como una colección de ítems o preguntas que miden algún domino o categoría (***factor,*** en términos estadísticos), y que, a su vez, conforman una condición (síndrome, constructo, variable) definida e identificada.



En otras palabras, una escala es un instrumento de medición consistente en un conjunto de preguntas (ítems) encaminadas a revelar determinados niveles de una o más categorías (dominios) que subyacen a un condición (síndrome, constructo), y que no podemos observar o medir directamente.

TIPOS DE INSTRUMENTOS

A pesar de que en la actualidad poco se considera el tipo de escala que debe utilizarse, es útil recordar el propósito para el cual fue creada cada una de ellas, y tomar en cuenta algunas de sus propiedades al momento de su adaptación

y su validación. En general, se puede decir que existen tres tipos de instrumentos[2]:

- Instrumentos ***discriminativos***.
- Instrumentos ***predictivos***.
- Instrumentos ***evaluativos***.

Instrumentos discriminativos

Son los utilizados para discriminar entre individuos respecto a una dimensión seleccionada. Su objetivo es describir un estado o una condición subyacentes, para lo cual no existe un patrón de oro: por ejemplo, diferenciar a los encuestados según la presencia o la ausencia de la condición medida o según la magnitud de presencia del aspecto evaluado. Las escalas que evalúan la inteligencia, la personalidad, la calidad de vida, el funcionamiento familiar o la adaptación a una enfermedad son ejemplos de estos instrumentos discriminativos. La utilidad de dichas escalas depende de la validez que tenga el constructo y de la confiabilidad o la repetibilidad de sus resultados.

Instrumentos predictivos

Son escalas que intentan predecir la pertenencia del sujeto evaluado a una o varias categorías definidas. Tratan de clasificar a los sujetos en las categorías de interés de acuerdo con un criterio externo disponible; esto es, un patrón de oro o un desenlace específico.

Los instrumentos más comunes dentro de esta categoría son las pruebas diagnósticas y los cuestionarios o entrevistas estructuradas, así como algunas escalas que buscan predecir algún desenlace (como la recaída en una enfermedad, la respuesta al tratamiento, etc.). La utilidad de dichos instrumentos depende de la confiabilidad y de la validez de criterio (exactitud en la clasificación de los sujetos).

Instrumentos evaluativos

Son los utilizados para evaluar la magnitud de cambio en el tiempo, en una dimensión o una categoría de interés. Estas escalas se usan, primariamente, para medir la evolución de una enfermedad o el resultado de un tratamiento. Como ejemplo de ello, la escala de Hamilton para depresión permite registrar en qué grado evoluciona la manifestación de la enfermedad tras un tratamiento psicoterapéutico o farmacológico; una escala de dolor sirve para evaluar los cambios que siente una persona después del uso de analgésicos, etc. La utilidad de este tipo de escalas depende de la validez de constructo, de la confiabilidad y de la sensibilidad al cambio o la capacidad de respuesta.

ADAPTACIÓN CULTURAL O VALIDACIÓN DE UNA ESCALA *VS.* CONSTRUCCIÓN Y VALIDACIÓN DE *NOVO*

Es común en los proyectos de investigación usar algún instrumento para medir una condición o una variable requeridas a la hora de obtener información fundamental para el estudio. A menudo, esta condición es el desenlace principal o una de las variables de interés. Una vez definido si se requiere o no el empleo de una escala, debe buscarse la que cumpla y desarrolle mejor tal cometido. Algunas veces, el instrumento existe, pero debe adaptarse culturalmente y validarse, para que sea legítimo utilizarlo en el medio y con los participantes en quienes se aplica.

En general, se requieren una adaptación y una validación cuando en el proceso se le hacen cambios importantes al instrumento original; también, cuando no se tiene, o no existe, la validación original; cuando se cambian la forma de aplicación, el idioma o el contenido, y cuando se quiere estar seguro de la comparabilidad con la versión original: por ejemplo, en un estudio multicéntrico, en el que se requiere revalidar las nuevas condiciones[3,4].

Por lo general, en muchos estudios, ya sean transculturales o no, se requiere el uso de instrumentos (escalas) que fueron desarrollados en otro ámbito y, a menudo, en otra cultura. Estos instrumentos algunas veces necesitan ser traducidos al idioma local, pero siempre necesitan una adaptación cultural, que puede incluir el cambio de términos y la revisión de conceptos y significados. Si no se hace, como se describirá más adelante, en el proceso de adaptación (que culmina, normalmente, en la validación de la escala), puede suceder que la validez y la confiabilidad originalmente establecidas para el sitio de origen del instrumento no se conserven, y su aplicación en el nuevo medio deje de ser adecuada. Por tal motivo, se debe aplicar una metodología que garantice la

reducción de sesgos culturales y maximice la utilidad de la escala[5-7].

Todo instrumento de medición, entre los que se cuentan las escalas utilizadas habitualmente en investigación, debe tener las siguientes características[8-12][1]: 1) validez, 2) confiabilidad, 3) adecuada amplitud de rango, 4) sensibilidad al cambio o capacidad de respuesta, y 5) utilidad práctica. El proceso que envuelve la verificación y realización de cada uno de ellos es lo que se llama validación de una escala.

Otras veces no existe la escala, o la que existe no es adecuada y debe construirse ***de novo***. En estos casos debe evaluarse si resulta útil emprender tan delicada actividad, que, además de ser compleja, requiere inversión de tiempo y de dinero. En general, se prefieren la adaptación y la validación de una escala, por cuanto hacerlo resulta más rápido y económico; se pueden utilizar instrumentos comúnmente aceptados, probados y utilizados, lo cual provee comparación con otros estudios, e, incluso, con otras personas, como en el caso de un estudio transcultural; además, requieren menor experiencia que el desarrollo ***de novo*** de un instrumento.

En la actualidad, existe una tercera opción, muy utilizada (pero no por ello la mejor opción en la mayoría de casos), denominada ***adaptación cultural***. En esta, una escala existente se somete al proceso de adaptación lingüística y cultural, no se le hace el proceso completo de validación y se asumen los criterios que se han establecido en su creación, aunque esto puede no ser cierto en la cultura en la cual se la aplica. En estos casos, a veces, se utilizan de manera un tanto amplia las características psicométricas que tiene originariamente.

Este proceso solo completa algunas de las etapas descritas para la validación: por ejemplo, establecer la traducción de los ítems, la traducción y la traducción inversa (retrotraducción), la validez de apariencia, la confiabilidad o algún otro proceso requerido; y con ello se utiliza el instrumento.

Este proceso de adaptación cultural puede resultar incompleto, pero en algunos casos no lo es, y puede ser un método más rápido y adecuado. No obstante, para que resulte completo es necesario tener datos objetivos que establezcan validaciones semejantes en culturas semejantes a la cultura donde se aplicará, o en un gran número de culturas diferentes; esto es, que el constructo, la condición o el síndrome que se esté midiendo y los resultados de la validación sean estables a lo largo de las diferentes culturas, lo cual no siempre es posible.

Por otra parte, puede ocurrir que la utilización repetida de la escala y sus resultados hayan sido coherentes con lo establecido, y lo esperado, entonces, es que el uso de la escala sea tan fácil y obvio (por ejemplo: una escala que mida la intensidad del dolor) que no requiera emprender todo el proceso, u otras razones que dependerán más del tipo de escala utilizada y del criterio del investigador.

No obstante lo anterior, si se espera que la escala varíe en medio de la aplicación, que los umbrales de "calificación" sean diferentes de los comúnmente establecidos por la apreciación que se tenga, por parte de los participantes, de algunos ítems o categorías; si se espera que el constructo o el síndrome se comporte de otra forma o tenga un número diferente de dominios; si las preguntas o los ítems no aplican, o requieren grandes cambios, es preferible hacer el proceso de adaptación y validación de la escala.

Pero, siendo estrictos, en caso de existir dudas sobre la validación original, de pensarse que el medio es diferente del que dio origen a la escala, de requerirse una medición precisa o de tener umbrales definidos, es preferible hacer todo el proceso de adaptación y validación. Es importante recordar que la mayoría de las veces no es adecuado extrapolar el proceso de validación de un instrumento generado en un ámbito cultural diferente del ámbito donde se aplicará.

PROCESO DE ADAPTACIÓN CULTURAL Y VALIDACIÓN DE ESCALAS

El proceso completo para la adaptación y validación de una escala consiste en las siguientes etapas[2] [4,8,13]:

a. **Búsqueda y selección del instrumento.**

[1] **Streiner DL. Scale construction checklist. Mc**-Master University Document. (Inédito).

[2] **Ver nota al pie 1.**

b. **Traducción y traducción en el sentido inverso** del instrumento.
c. **Prueba piloto para evaluar ajuste del ítem,** utilidad y formatos.
d. **Evaluación de validez.**
e. **Evaluación de confiabilidad.**
f. **Evaluación de sensibilidad al cambio o a la** capacidad de respuesta.

Búsqueda y selección del instrumento

Una vez determinada la pregunta de la investigación, a corto plazo surgen cuestionamientos acerca de las variables, las condiciones y los desenlaces que se van a evaluar, sobre cómo se medirán y con qué tipo de instrumentos se hará la medición.

Algunas veces se tienen escalas pertinentes y validadas, aunque en la mayoría de los casos las escalas útiles para investigación en ciencias de la salud están en otro idioma o no son directamente aplicables. Y a menudo no hay escalas disponibles que sean útiles para la medición de interés.

Una vez se han determinado las condiciones y las variables en un estudio, se debe definir cómo se van a medir. Para ello se deben usar cuestionarios, pruebas de laboratorio, imágenes diagnósticas, entrevistas y, en ocasiones, escalas, entre otros recursos. Posteriormente, se debe definir cuál es el instrumento más adecuado psicométricamente y por uso, el cual suele ser el más comúnmente reportado en la literatura en estudios similares para medir la condición o la variable subyacentes. En los casos en los que hay varias opciones disponibles, debe preferirse la más utilizada en el medio para el tipo de estudio concebido, lo cual mejora la calidad de los resultados que se van a comparar.

Existen hoy día un sinnúmero de instrumentos; algunos de ellos son de uso libre y otros requieren pagos. Debe indagarse sobre sus condiciones de uso (habitualmente, en sitios dedicados al tema en Internet). Así mismo, existen libros dedicados a la descripción de instrumentos, con resúmenes de sus características psicométricas.

Por otra parte, y con el fin de saber más acerca de utilización de estos instrumentos, se puede hacer una búsqueda en diferentes sitios de internet que proveen información sobre los artículos, como Medline, Pubmed, Lilacs, SciELO, PsycINFO, entre otras bases. Posiblemente, dado que las escalas son de común uso en ciencias sociales, PsycINFO provee una información más completa sobre los artículos que las han usado, pero en el caso de la investigación biomédica, las dos primeras pueden proveer información de importancia. Este proceso debe llevar a una cuidadosa revisión de la literatura, y en algunos casos, a preguntar a expertos en el tema sobre la mejor escala disponible.

Una vez concluye la búsqueda de las escalas de medición comúnmente utilizadas, se evalúa con sentido crítico su calidad, tomando en cuenta el diseño del estudio utilizado en su desarrollo. Deben evaluarse, entre otros aspectos:

- Credibilidad y aplicabilidad cultural de la teoría subyacente.
- Tipo de instrumento y de adecuación para el estudio (predictivo, discriminativo o evaluativo).
- Diseño y aplicabilidad de los ítems al medio.
- Tipo de respuestas escalares y su aplicabilidad al medio (véase el **capítulo 28** "Desarrollo de cuestionarios").
- Selección y completitud de ítems.
- Posibles sesgos culturales en la aplicación de las escalas.
- Confiabilidad, validez y sensibilidad al cambio[3(13,14)].

En algunos casos, se encuentran defectos o problemas tras la evaluación del instrumento original, y que al ser traducidos y adaptados culturalmente podrían ser magnificados. En estos casos, definitivamente, es más apropiado crear un nuevo instrumento.

Cabe recordar que desarrollar o construir una nueva escala a partir de la población de estudio y de la propia experiencia es una opción que consume mucho tiempo y dinero; por lo tanto, en términos generales, es preferible, en cuanto sea posible, utilizar una escala que exista; aún mejor, si ha sido adaptada y validada en el medio.

En caso de existir una escala no adaptada o no validada, el investigador debe continuar con el siguiente paso; o sea, traducir y adaptar

3 **Bowman J, Sanson-Fisher R. Questionnaire construction:** An Introductory Guide to Questionnaire Construction. (Inédito). 1987.

culturalmente la escala seleccionada. Así mismo, si se prevé que se culminará con la validación completa de la escala es recomendable que el investigador procure tener contacto con el autor de la escala, para saber si se está llevando a cabo su validación en el medio de aplicación, o si no es así, para que oficialmente se le autorice realizarla, cuestión que es conveniente para las dos partes.

Traducción y traducción en sentido inverso

El instrumento elegido para el estudio deberá ser revisado por un comité de traducción compuesto por tres o cuatro personas, dentro de las cuales debe haber, por lo menos, una persona con experiencia en el tema, en la aplicación del instrumento y con buen manejo del idioma original, así como una a dos personas que presenten o hayan presentado la condición medida, e, idealmente, conocedores o hablantes del idioma original. Este grupo dará sus recomendaciones sobre el instrumento, y, ante todo, debe anticipar posibles problemas en los ítems. Por ejemplo, si se pregunta el nombre del Rey de España, podría ser cambiado por el nombre del Presidente del país; o si se indaga acerca de determinados alimentos, como carne de búfalo o cítricos, reemplazarlos por carne de res o naranjas, respectivamente; o si se averigua acerca de aficiones como el cricket o el hockey, reemplazarlas por dominó y fútbol.

Una vez hecha esta revisión y sugeridos esos cambios, se recomienda iniciar el proceso de traducir la versión en su idioma original al español por parte de dos traductores, que deben ser personas calificadas; esto es, bilingües o con buen dominio del idioma original (inglés, alemán, francés, etc.), y que tengan un buen nivel de habla nativa en español; además, deben vivir en el país del caso y conocer muy bien la cultura con la cual se trabajará, conocer las características mínimas de instrumentos y de medición en salud (si se requiere, los traductores podrían ser capacitados por el investigador principal) y algún conocimiento del área específica (psiquiatría, neumología, etc.).

Cabe anotar que, ante todo, se requiere una buena traducción conceptual y lingüística del instrumento original. Estas primeras versiones traducidas al español serán evaluadas y conciliadas por el comité de traducción, cuyos miembros escogerán la versión más adecuada o conciliarán y ensamblarán una versión derivada de las dos obtenidas. Para esta conciliación puede requerirse, si hay diferencias importantes, que los traductores se reúnan y concilien sobre sus diferencias. No obstante, por lo general, eso no es necesario.

Posteriormente, la versión en español será traducida al idioma original (inglés, alemán, etc.) por dos personas bilingües o nativas en el idioma de origen y con buen dominio del español, además de algún conocimiento de la cultura que le dio origen y el área específica de interés del instrumento. Ninguno de los traductores debe conocer la fuente en el idioma original. Las versiones obtenidas en la traducción inversa serán revisadas por el comité, y así se obtendrá una versión final, que será comparada con la original. En caso de que las dos versiones sean iguales, esta etapa habrá culminado; si hay diferencias significativas (que puedan alterar los resultados o la comprensión), es recomendable conciliar las versiones con los traductores, y, en último caso, enviar la nueva versión al autor de la escala, con el fin de disponer de sus sugerencias y sus recomendaciones como experto en la materia.

Es importante recordar que esta etapa tiene por objeto lograr lo que algunos autores[15] han llamado "equivalencia translingüística" entre el instrumento original y el traducido, y que esta hace parte fundamental de la adaptación cultural del instrumento. El instrumento traducido obtenido pasará a la siguiente etapa, que consiste en una prueba piloto para su evaluación preliminar.

Prueba piloto: evaluación y ajuste de ítems; utilidad y formato del instrumento

Antes de evaluar el instrumento en un estudio formal de validación, es recomendable hacer una prueba piloto, para probar su uso, refinar los aspectos técnicos (tamaños de espacios, legibilidad, portabilidad) y mejorar y corregir todas las deficiencias que puedan encontrarse. En general, se pretende que la prueba brinde:

- Evaluación de los ítems.
- Evaluación de la utilidad de la escala.
- Evaluación del formato del instrumento.

En la prueba piloto se aplica el instrumento

a 10 o 20 personas que tengan el mismo nivel educativo y, en lo posible, la misma enfermedad que las personas a quienes se les piensa aplicar, teniendo presente que el instrumento debe dar cuenta de los diferentes espectros de intensidad de la patología o la variedad en la cual se presenta. Durante la aplicación, el investigador debe hallarse presente, evaluar cómo responden los sujetos y buscar la interpretación de cada punto. En caso de que se presenten dudas, conviene indagar acerca de las diferentes interpretaciones que se le dan al ítem (por ejemplo, pedir al sujeto que exprese con sus propias palabras qué entendió), averiguar por dificultades específicas con cada una de las preguntas y precisar cuáles de ellas resultan confusas o poco claras.

Evaluación de los ítems

Algunos autores[5,16] describen dos métodos para clarificar puntos en cuestionarios autoadministrados:

a. **Después de que ha contestado un ítem, se le** pide al sujeto que califique, en una escala de tres puntos (entre fácil, moderada y difícil), la dificultad de cada pregunta.
b. **Se le pide al sujeto subrayar las palabras que** no entendió del punto o sobre las cuales tiene dudas.

Si la mayoría de los sujetos evalúan la pregunta (ítem) de la misma forma, esta pregunta deberá ser revisada, pues, muy probablemente, no discriminará a los encuestados, o su validez podría hallarse en entredicho.

Por otra parte, en la evaluación general de los ítems (preguntas) se deben verificar los siguientes aspectos[1]:

- Comprensión: **En general, se debe procurar** que la escala sea comprensible para personas con cualquier nivel de educación. Para ello, es deseable buscar un nivel que sea comprensible para un niño de 8-10 años. Así mismo, los términos deben ser sencillos: se debe evitar el uso de una terminología técnica, compleja o de uso poco común.
- Evitar ambigüedad del ítem: **Se debe evitar el** uso de palabras que puedan tener diferentes interpretaciones. Por ejemplo, usar "a menudo" puede significar dos, tres, cinco o más veces a la semana, según la persona y el tipo de variable investigada. En estos casos debe evitarse el uso de esas palabras, o tratar de ser más preciso en su interpretación. Ante esta situación, sería deseable especificar cuántas veces por semana o preguntar directamente *qué tan a menudo en el día*: **una, dos, tres** o más veces. No tener en cuenta esto puede generar errores de medición.
- Buena redacción del ítem: **Se busca disminuir los** aspectos que confundan (dobles negativos); por ejemplo: "Usted no estaría de acuerdo con que no se acepte la migración de las personas"; también, evitar las palabras con carga afectiva; por ejemplo: "Habitualmente, usted acepta las pésimas decisiones de su jefe", y la redacción confusa o compleja, entre otras posibilidades (para mayor información, **(véase el capítulo 28 "Desarrollo de** cuestionarios").
- Rango de respuesta: **Hay una tendencia a** evitar los extremos de las opciones de respuesta (***sesgo de aversión a los extremos***) que debe evaluarse, y dejar más niveles si se requiere una mayor diversidad de respuesta. Lo ideal, y muy utilizado, es tener respuestas en todo el rango requerido. Por otra parte, debe considerarse lo contrario si, por ejemplo, se tienen cinco opciones de respuesta: "Muy Malo", "Malo", "Regular", "Bueno" y "Muy bueno"; pero si en el piloto no se obtienen respuestas de muy malo, debería considerarse la necesidad de retirar dicha opción y dejar solo cuatro niveles, siempre y cuando el tamaño de la muestra piloto represente los diferentes espectros de intensidad y de variabilidad.
- Adecuada frecuencia de respuesta: **Si una pre**gunta tiene la misma respuesta en más del 95 % de los casos, debe ser retirada, por su falta de discriminación del ítem y su poco aporte a la variabilidad de respuesta al instrumento. Sin embargo, como esta prueba piloto se hace con pocas personas, ello puede corresponder a un error de medición, y ante la duda se debe dejar en observación el ítem durante las otras etapas de la validación, con el fin de determinar si se excluye o no de la escala.

Utilidad del instrumento

Consiste en que la escala o el instrumento de medición se pueda utilizar y aplicar fácilmente en la vida real. Si una escala es tan costosa que solo se puede aplicar en centros asistenciales muy específicos y exclusivos, no será muy útil

en el medio general. También tendrá poca utilidad, aunque sea fácil de adquirir, pero difícil de aplicar, o si es muy dispendiosa o requiere un entrenamiento muy complejo para manipularla. Algunos de los aspectos que tienen que ver con este punto son:

- Tiempo de aplicación y de diligenciamiento del instrumento: En los dos casos, si el instrumento es aplicado por un tercero o es autodiligenciado, se requiere medir el tiempo necesario para ello. Sin embargo, el instrumento debe ser suficiente y razonable. Las escalas muy largas disminuyen la colaboración de los sujetos y pueden fatigar al propio encuestado; además, aumentan los costos y disminuyen la confiabilidad y la validez del instrumento. Como regla general, cuanto menos tiempo exija, más sencilla sea y menos ítems tenga la escala, tanto mejor. Teniendo en cuenta que a menor tiempo mejor es el instrumento, se tiene una mayor posibilidad de que este sea aceptado y respondido con validez.
- Necesidad de entrenamiento: Cuanto menor sea el entrenamiento necesario para la aplicación del instrumento, mayor será su facilidad de difusión, y mayor y más económica su aplicación.
- Facilidad para la calificación: Cuantos menos elementos sofisticados se requieran para su calificación (ponderación de ítems, algoritmos, computaciones, etc.), mayor será su uso, y mejor su aplicación.

Formato del instrumento

En esta etapa es útil evaluar:

- Características del formato: El formato, en general, debe ser *limpio* (con pocos datos, o solo los necesarios), con un instructivo corto y completo para lo requerido, con claridad acerca de los ítems y de los tipos de respuesta, con un tipo de letra atractivo y fácil de leer para las personas que lo van a responder (por lo general, letra de tamaño 12). No es conveniente tener todas las letras en mayúsculas, por cuanto esto demora su lectura.

Para hacer una evaluación del instrumento, es útil revisar si se han cumplido las recomendaciones. Se sugiere tener una lista de revisión, que permita mejorar su calidad. Las siguientes preguntas han sido propuestas como medio para evaluar el instrumento como último paso:

- ¿Son claras las instrucciones?
- ¿Es conveniente poner un ejemplo en las instrucciones?
- ¿El método usado para obtener las respuestas es claro?
- ¿Cuán fácil es evitar (esquivar) una pregunta?
- ¿Se requieren saltos entre algunas preguntas?
- ¿Sería útil un formato electrónico (computarizado) para su uso?
- ¿Qué otras dificultades se pueden prever con el formato utilizado?

EVALUACIÓN DE VALIDEZ



Se refiere a que la escala o el instrumento mida lo que dice medir; o sea, que el resultado obtenido corresponda al fenómeno real, y no a otro: por ejemplo, un instrumento para medir la ansiedad no debe producir resultados positivos en pacientes que solo tienen un trastorno conversivo.

Dentro de los tipos de validez tenemos las siguientes: 1) de apariencia, 2) de contenido, 3) de criterio y 5) de constructo.

Validez de apariencia

Se obtiene mediante la evaluación del instrumento por dos grupos de jueces, expertos y pacientes o personas con la condición estudiada, de cuatro a cinco personas cada uno, y que dicen si, en su concepto, el instrumento parece medir las cualidades deseadas. O, dicho de otra manera, responden a la pregunta: *¿La escala parece medir lo que debe medir?* Esto es, cada grupo evalúa si la condición se mide, si los ítems, en apariencia, representan todos los dominios, las categorías o los factores (en términos estadísticos) subyacentes a la condición estudiada. En caso de que la respuesta sea afirmativa, provee aceptabilidad al momento de aplicarla. No obstante, en esta parte no se hace ninguna prueba estadística, sino, simplemente, se provee un concepto, el cual, en ocasiones, es tan sencillo como que "parece que mide la condición X, pero, profundizando un poco más, se podría decir que parece medir la condición X y todas las dimensiones (dominios) que, se supone o se cree teóricamente, debería contener".

Este tipo de validez es tan solo el primer paso hacia los siguientes, aunque, en ocasiones, es el único que se hace no solo al "validar" la escala, sino al construirla por primera vez. De ello da cuenta la dificultad de encontrar los desarrollos psicométricos o la validación de una determinada escala.

Validez de contenido

Responde a la pregunta: *¿La escala mide todos los dominios o los factores subyacentes que corresponden a la condición (concepto, síndrome o constructo) estudiada?* Es un concepto relacionado con el anterior, pero implica un trabajo matemático (análisis factorial), en el que se evalúa si el instrumento explora todas las categorías (dimensiones) o dominios pertinentes a la condición, el concepto o el constructo en estudio.

Un ejemplo de lo anterior sería el síndrome depresivo, que puede clasificar varios trastornos, uno de los cuales es el trastorno depresivo recurrente (o trastorno depresivo mayor). Si se quiere medir dicho trastorno mediante una escala, se podría pensar en varios síntomas que se constituirían en ítem, tales como: astenia, anorexia, adinamia, tristeza, llanto, dificultad en la concentración, quejas de memoria, hipomimia, irritabilidad, agitación o inhibición, etc., y que van a depender de la persona que los aqueja. Estos síntomas y signos se organizan en preguntas o ítems, que, a su vez, podrían agruparse en una serie de dominios o categorías (***factores***, en el lenguaje estadístico) mayores, como: estado de ánimo; función motriz; función autonómica, función cognitiva y, para algunos, la función social, o de relación.

De esta manera, el trastorno depresivo se podría construir teniendo en cuenta alteraciones de diversa índole (síntomas o signos que se reflejan en ítems) constitutivos y representativos de estos cinco dominios (categorías) o factores subyacentes. Esto es lo que se llama ***estructura de factores o dominios*** **(tabla 8.1)**.

En resumen, la validez de contenido hace referencia a la capacidad del instrumento para contener y medir todas las categorías o los dominios estrechamente relacionados con el fenómeno o la condición que se estudia. Por ejemplo, una escala para trastorno obsesivo debe medir tanto el espectro obsesivo (primer domino) como el compulsivo (segundo dominio). Una escala de calidad de vida debe medir, mediante ítems relevantes, todas las dimensiones que componen dicho constructo, tales como: funcionalidades física, emocional y social; ejecución de su papel en la vida; dolor y otras categorías pertinentes, de acuerdo con la escala. Esto es lo que se denomina ***amplitud de rango***, pues debe contener todos los factores o los dominios pertinentes.

Ahora, para garantizar la validez de contenido se deben utilizar métodos estadísticos, entre los cuales el ***análisis factorial*** (AF) es el más ampliamente utilizado[17-19]. Dicho análisis procura determinar qué factores (dominios) subyacen a la condición estudiada, y, a la vez, qué ítems los representan. Por lo tanto, el AF ayuda a determinar cuál es la estructura factorial subyacente a la condición (síndrome, constructo) estudiada, y, a la vez, cuáles ítems son los más adecuados para representar cada dominio; es decir, cuáles preguntas ayudan a explicar la mayor variabilidad posible, así como a determinar si dichas preguntas siguen una estructura específica de uno o de varios factores (en caso de que existan). A la vez, permite decidir sobre cuál o cuáles ítems, definitivamente, se podrían retirar del constructo por su baja contribución a los nuevos factores, lo cual lleva a que sean redundantes y aporten poco a la variabilidad explicada.



En términos generales, puede decirse que el AF tiene como objetivos: 1) reducir el número de variables; 2) evaluar la posible presencia de multicolinealidad entre las variables correlacionadas; 3) evaluar y detectar la unidimensionalidad o, al contrario, la multidimensionalidad de los constructos; 4) evaluar la validez del constructo; 5) examinar la relación o la estructura de las variables; 6) desarrollar constructos teóricos, y 7) probar teorías propuestas[20,21].

Existen dos tipos principales de AF: 1) el confirmatorio (AFC) y 2) el exploratorio (AFE). Al respecto, hay cierta controversia sobre cuál utilizar en la validación de una escala. Muchos autores en ciencias sociales tienden a preferir el uso del AFE como primera opción para investigar la estructura factorial subyacente y refinar el *pool* de ítems que se tenga. Se propone, a la vez, que el AFE debe continuarse por un AFC usando muestras diferentes, planteando ***a priori*** el modelo hipotetizado y los modelos competitivos que se tengan.

Tabla 8.1 Síndrome depresivo: dominios e ítems que lo pueden representar

Categoría, dominio o factor del síndrome	Ítems que podrían representar cada dominio (factor)
Estado de ánimo	**Durante los últimos 14 días:** • ¿He estado triste? • ¿He disfrutado lo que hago? • ¿He sentido placer al hacer las cosas? • ¿No he tenido deseos de hacer nada? • ¿He llorado todos los días?
Función motriz	**Durante los últimos 14 días:** • ¿Se me ha dificultado moverme? • ¿Me han pesado los brazos y las piernas? • ¿He hecho con lentitud las cosas? • ¿Me he fatigado más que de costumbre al hacer las cosas? • ¿He presentado agitación o inquietud motriz? • ¿En las tardes me siento sin deseos de hacer nada?
Función autonómica	**Durante los últimos 14 días:** • ¿He tenido más apetito que de costumbre? • ¿Mi apetito ha disminuido? • ¿No he tenido apetito? • ¿Se me ha dificultado quedarme dormido (insomnio de conciliación)? • ¿Me he despertado en la madrugada (insomnio de reconciliación)?
Función cognitiva	**Durante los últimos 14 días:** • ¿Me he sentido culpable? • ¿Mi pensamiento ha estado lento? • ¿He tenido dificultad para pensar? • ¿He tenido ideas suicidas?
Función social o de relación	**Durante los últimos 14 días:** • ¿Me he irritado fácilmente con otras personas? • ¿No me ha interesado ver a nadie? • ¿He querido estar solo? • ¿He pensado que nadie es interesante? • ¿No he sentido deseos de estar dialogando con alguien?

Fuente: Elaboración de los autores.

En el caso del AFE, por tratarse de una técnica de interdependencia, lo que se busca es identificar una estructura interna al generar nuevos factores. Para ello, es necesario considerar algunos pasos básicos, como verificar los supuestos y establecer el método de derivación de los factores, así como la evaluación y la interpretación de los factores[19].

Respecto a los supuestos, ellos se refieren a si tiene sentido hacer un AF; por lo tanto, la recomendación principal es evaluar si existe correlación entre los ítems o las variables, ya que en caso de no existir, no se justifica su uso. Para hacer esta evaluación, se debe calcular la matriz de correlaciones, así como algunas de las siguientes pruebas dedicadas para tal fin: 1) la prueba de esfericidad de Bartlett; 2) el índice Kaiser-Meyer-Olkin (KMO) y 3) el índice de adecuación de la muestra individual (MSA). De manera sucinta, en la **tabla 8.2** se presentan los criterios para tomar esta clase de decisión.

A manera de ejemplo, se van a utilizar los resultados obtenidos en un estudio donde se buscaba determinar, en un grupo de 267 pacientes, la estructura factorial al aplicarles la escala de depresión de Zung, la cual se basa en 20 ítems.

Como primer paso, se evalúa si tiene sentido hacer una AFE; por lo tanto, se calcula la respectiva matriz de correlación, así como las pruebas y los índices ya mencionados. Al calcular la matriz de correlación, uno de los problemas que se presentan es que a medida que el número de ítems se incrementa, se hace más difícil su visualización, así como la posibilidad de que al menos en 2 ítems haya correlación se incrementa. En el ejemplo, dado que se trata de 20 ítems, la matriz es bastante amplia, así que, a manera de ilustración **(tabla 8.3)**, se va a mostrar la correlación que hay entre los primeros 5,

Tabla 8.2 Descripción de las pruebas y los criterios para evaluar si tiene sentido hacer un AFE

Prueba	Descripción	Evaluación	Decisión
Matriz de correlación	**Matriz que contiene las correlaciones** simples entre cada uno de los ítems. Toma valores entre 0 y 1.	Correlación entre: • |0,000-0,299|: baja. • |0,300-0,699|: moderada. • |0,700-1,000|: alta.	Si todas las correlaciones son bajas, no tiene sentido hacer el **AF.**
Prueba de Esfericidad de Bartlett	Esta prueba se basa en la matriz de correlación, y plantea como hipótesis nula que no hay correlación entre las variables.	Utiliza una prueba de χ^2 para su cálculo.	**Hacer el AF si se rechaza la** hipótesis nula.
Índice KMO	Se basa en el cálculo de correlaciones parciales entre todos los ítems. Al igual que las correlaciones simples, toma valores entre 0 y 1.	Si el valor del índice es: • < 0,5: inaceptable. • 0,5-0,59: pobre. • 0,6-0,79: regular. • 0,8-1: meritorio.	Hacer el AF si se encuentran valores regulares o meritorios.
Índice MSA	A diferencia del KMO, calcula las correlaciones parciales considerando solo uno de los ítems. Al igual que las correlaciones simples, toma valores entre 0 y 1.	Si el valor del índice es: • **< 0,5: inaceptable.** • 0,5 a 0,59: pobre. • 0,6 a 0,79: regular. • 0,8 a 1: meritorio.	Hacer el AF si se encuentran valores regulares o meritorios.

Fuente: Elaboración de los autores.

donde puede verse una correlación moderada entre algunos de ellos.

Al estimar la prueba de esfericidad de Bartlett, se encontró un valor de $p < 0{,}001$, lo cual indica que se rechaza la hipótesis nula. El índice KMO fue de 0,89, en tanto que el valor más pequeño del índice MSA fue el del ítem 7, con un valor de 0,77. Al evaluar estos cuatro resultados, se puede decir que sí tiene sentido hacer un AFE.

Una vez definido dicho aspecto, se debe establecer cuál será el método de derivación de los factores, así como la evaluación del ajuste. En ese sentido, es importante definir *a priori* el método por usar para extraer los nuevos factores (los más recomendados son el método de factores comunes y el método de factores por componentes principales[22]), y con qué criterio se va decidir cuántos factores se deben seleccionar para el análisis (tres de los más utilizados son: 1) el criterio del test de pendiente (*scree test*), 2) el criterio del valor propio (*eigenvalue*), y 3) el criterio del porcentaje de varianza explicada).

Tabla 8.3 Matriz de correlación de los primeros cinco ítems de la escala de Zung

Ítem	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5
Ítem 1	1,00				
Ítem 2	- 0,24	1,00			
Ítem 3	0,60	- 0,26	1,00		
Ítem 4	0,44	- 0,20	0,51	1,00	
Ítem 5	- 0,27	0,17	- 0,32	- 0,26	1,00

Fuente: Elaboración de los autores.

Los resultados obtenidos al hacer el análisis por el método de factores comunes se presentan en la **tabla 8.4**. A partir de estos resultados, se procede a decidir cuál será el número de factores que se van a retener. Una primera aproximación se hace a través del *scree plot* **(figura 8.1)**, donde se busca en el gráfico el punto de inflexión en el cual la tendencia de la línea cambie de concavidad, y ese punto corresponde al valor propio del máximo factor para retener. En este caso, corresponde al factor número 5. Sin embargo, como se puede apreciar, es un cambio muy sutil; muchas veces, poco detectable para el investigador, lo cual la hace una estrategia un tanto subjetiva. Los mismos resultados se obtienen al utilizar la estrategia de evaluar los valores propios > 1.

En cuanto al tercer criterio, este queda a discreción del investigador, quien decide de antemano qué porcentaje de varianza común quiere explicar con sus nuevos factores. En este caso, si se supone que se quiere explicar el 70 % de varianza, se necesitaría seleccionar para el análisis un total de 8 factores. Como este criterio no se estableció de antemano, el análisis se seguirá haciendo con 5 factores.

Tabla 8.4 Descripción de los valores propios y los porcentajes de varianza explicada de los nuevos factores estimados a partir del método de factores comunes

Factor	Valor propio	% de varianza explicada	% acumulado
Factor 1	6,48	32,40	32,40
Factor 2	1,77	8,85	41,25
Factor 3	1,42	7,12	48,37
Factor 4	1,11	5,54	53,91
Factor 5	1,05	5,25	59,16
Factor 6	0,92	4,58	63,74
Factor 7	0,86	4,29	68,03
Factor 8	0,74	3,69	71,72
Factor 9	0,70	3,51	75,23
Factor 10	0,67	3,37	78,60
Factor 11	0,63	3,16	81,76
Factor 12	0,58	2,88	84,64
Factor 13	0,52	2,62	87,26
Factor 14	0,45	2,26	89,52
Factor 15	0,42	2,11	91,63
Factor 16	0,40	2,01	93,64
Factor 17	0,35	1,75	95,39
Factor 18	0,34	1,68	97,07
Factor 19	0,30	1,50	98,57
Factor 20	0,28	1,43	100,00

Fuente: Elaboración de los autores.

Una vez llevada a cabo esta selección, se deben hacer la evaluación y la interpretación respectivas. Para una mejor interpretación, es recomendable hacer una rotación de los nuevos factores. Sin embargo, esta rotación dependerá de la experiencia y el conocimiento previo que el investigador tenga del problema.

Si los nuevos factores no están correlacionados entre sí, se recomienda hacer una rotación ortogonal (independencia de los factores), dentro de las cuales se contemplan tres estrategias: 1) ***varimax*** (busca que cada ítem solo sea representativo en uno de los factores), 2) ***quartimax*** (busca minimizar el número de factores) y 3) ***equamax*** (es una combinación de las dos estrategias anteriores)[19,22,23].

En caso de que los nuevos factores estén correlacionados, la rotación recomendada es la oblicua. Las estrategias más usadas son 1) ***oblimin*** y 2) ***promax***. En el caso de la rotación *oblimin*, se debe establecer el grado de inclinación (δ) entre ellos, mientras la rotación ***promax*** transforma una rotación ortogonal hasta llegar a una solución similar a la "estructura ideal". En estos dos últimos escenarios, para evitar la manipulación de los resultados por parte del investigador, se recomienda establecer los criterios antes de recolectar la información.

Continuando con el ejemplo, dado que no se establecieron de antemano criterios de correlación entre los nuevos factores, se decidió hacer una rotación ortogonal para maximizar el

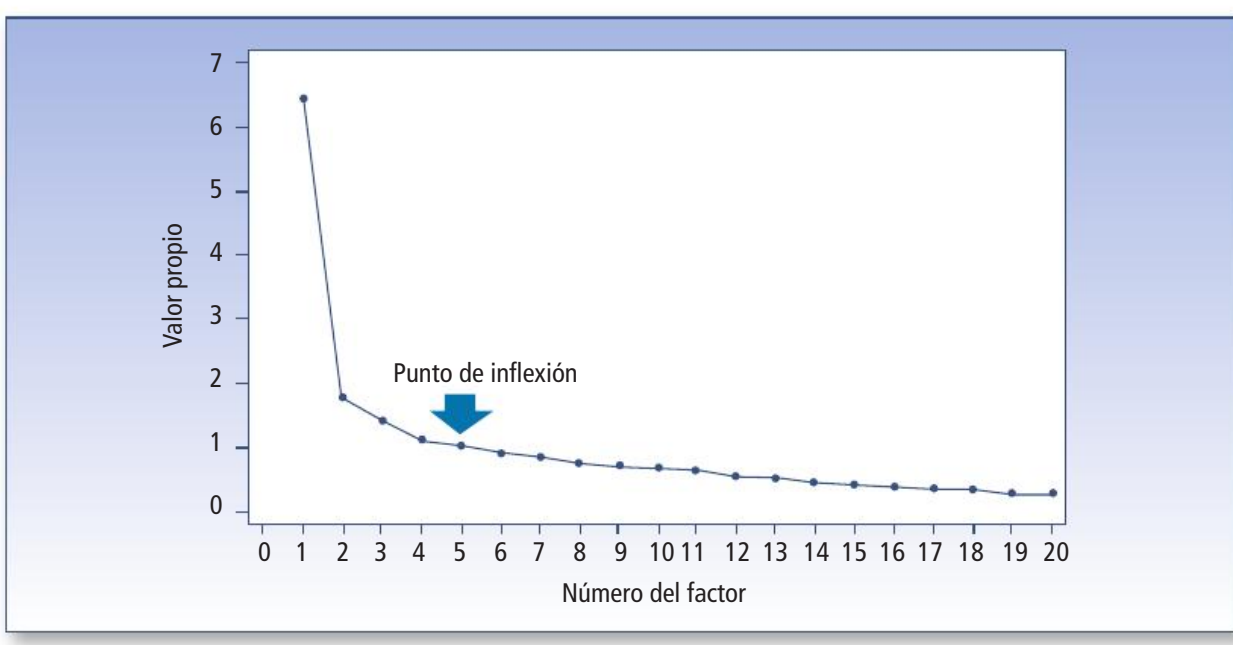


Figura 8.1 Criterio del test de pendiente de los valores propios, y porcentajes de varianza explicada de los nuevos factores.
Fuente: Elaboración de los autores.

peso de los ítems a uno de los criterios. La tabla 8.5 presenta los resultados obtenidos al realizar esta rotación. Para efectos de tener una mejor comprensión de los resultados, se eliminaron de la tabla las cargas que tuviera valores < 0,40, que, se podría decir, son bajas.

También puede verse que todos los ítems quedaron representados en algunos de los factores. No obstante, tres de ellos (ítem 4, ítem 11 e ítem 12), quedaron representados en dos factores a la vez. Es más: curiosamente, dos de ellos (ítem 4 e ítem 11) presentaron la misma ponderación en los dos factores, lo cual indica que ese ítem puede ser considerado en alguno de los dos o en ambos. En el caso del ítem 12, claramente se ve una mayor contribución en el factor 4.

En la columna derecha de la tabla 8.5, aparece la *comunalidad*, que representa la proporción explicada de cada uno de los ítems originales con la nueva estructura factorial; por lo tanto, los ítems 2, 5 y 13 están poco representados, y ello indica que su aporte final a la nueva estructura factorial es muy bajo, mientras que, por el contrario, el ítem 1 es el que se encuentra mejor explicado en esta nueva estructura (70 %).

Una vez realizado este proceso, sigue uno, quizá, más importante y complicado: tratar de ver qué tipo de relación hay entre los ítems de cada uno de los nuevos factores, y, si es posible, asignarles nombres lógicos a los nuevos factores que identifiquen qué representan en la estructura. En ese orden de ideas, la tabla 8.5 presenta cómo quedaron formados finalmente los nuevos factores.

En resumen, Cabrera-Nguyen[23] sugieren utilizar para reportar AFE:

- Una estrategia de rotación oblicua en el AFE, puesto que en ciencias sociales las correlaciones entre factores son la norma.
- La rotación oblicua y la rotación ortogonal proveen resultados similares si los factores no están correlacionados
- Si se tienen pocos parámetros, la rotación ortogonal puede ser más parsimoniosa.
- Con el fin de retirar ítems, se sugiere utilizar como criterio retirar los que carguen más de 0,32 en dos factores; especialmente, si hay ítems que tengan cargas > 0,5. Una vez excluido algún ítem, se debe correr de nuevo el AFE.

Para reportar los factores, deben tenerse en cuenta múltiples factores: considere valores propios (*eigenvalues*) > 1; rechace un factor con menos de tres ítems; tenga en cuenta el análisis paralelo[24] y el *scree test*[25].

En caso de optar por un AFC, se debe evitar el uso de análisis por componentes principales para comenzar.

Si se utiliza el AFC, se debe explicar por qué se opta por este tipo de análisis, y proporcionar la estructura factorial hipotetizada.

Finalmente, respecto al mínimo tamaño de la muestra necesario para un AF, hay una gran diversidad de opiniones al respecto. Algunos autores consideran que se requiere tener entre 5 y 10 personas por cada ítem estudiado, en tanto otros consideran que el tamaño de muestra mínimo debería ser de 100 personas con diferentes espectros de intensidad, y de 300 personas para un estudio de validación de escala. En su guía, Comrey[26] clasificó de la siguiente manera los tamaños de la muestra: 100, como *pobre*; 200, como *razonable*; 300, como *bueno*; 500, como *muy bueno*, y 1000 o más, como *excelente*.

Por otra parte, algunos autores[22,25] dicen que realmente se debe procurar la mayor muestra posible, a sabiendas de que solo se sabrá cuál era la muestra adecuada después de correr los análisis factoriales.



Sin embargo, existen dos reglas simples para la selección de la muestra que se deberían aplicar en este tipo de estudios: "Con 10 personas por ítem estudiado y por lo menos 100 personas en la muestra"[27,28].

No debe olvidarse que la muestra depende, así mismo, qué tanto cargan los factores, porque si dicha carga es de 0,4, una muestra de 150-200 podría resultar adecuada, e, incluso, si las cargas son > 0,8, un tamaño de muestra de 50 ya puede ser suficiente[23,25].

Validez de criterio



Este tipo de validez se aplica cuando existe un patrón de oro o patrón de referencia (instrumento, entrevista, etc.) y el instrumento en evaluación produce resultados que concuerden con dicho patrón.

Tabla 8.5 Distribución de las cargas factoriales en cada uno de los nuevos factores luego de hacer la rotación ortogonal *varimax*

Ítem	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Comunalidad
Ítem 1		0,77				0,70
Ítem 2				0,48		0,43
Ítem 3		0,66				0,61
Ítem 4		0,46			0,46	0,57
Ítem 5				0,46		0,43
Ítem 6				0,71		0,66
Ítem 7					0,80	0,66
Ítem 8			0,78			0,65
Ítem 9			0,76			0,67
Ítem 10			0,47			0,57
Ítem 11	0,50			0,50		0,62
Ítem 12	0,46			0,62		0,63
Ítem 13			0,44			0,44
Ítem 14	0,76					0,62
Ítem 15		0,56				0,50
Ítem 16	0,62					0,54
Ítem 17	0,79					0,66
Ítem 18	0,76					0,65
Ítem 19		0,68				0,54
Ítem 20	0,68					0,66

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Tiene la mente tan despejada como siempre.	Se siente triste y decaído.	**Tiene dificultad para** hacer deposición.	**Se siente mejor en las** mañanas.	**Tiene dificultades** para dormir.
Tiene confianza en el futuro.	**Siente ganas de llorar o** tiene periodos de llanto.	**Tiene palpitaciones.**	**Come igual que siempre** ha comido.	**Nota que ha** perdido peso.
Encuentra fácil tomar decisiones.	**Tiene dificultades para** dormir.	**Se siente cansado sin** razón aparente.	**Disfruta el sexo como** antes.	
Se siente útil y necesario.	Está más irritable que de costumbre.	**Se siente intranquilo;** no puede quedarse quieto.	Tiene la mente tan despejada como siempre.	
Encuentra agradable su vida.	**Cree que otros estarían** mejor si usted muriera.		**Encuentra fácil hacer las** cosas que hacía antes.	
Disfruta ahora de sus actividades cotidianas.				

Fuente: Elaboración de los autores.

Esto es, que la escala funcione de forma parecida a la de otros instrumentos. La validez de criterio provee, justamente, un criterio empírico de validez, y se divide en dos:

- Validez concurrente: Hace referencia al establecimiento de correlación entre el nuevo instrumento y una escala que ya ha sido probada (medida externa de criterio, patrón de oro u

otro). Por ejemplo, una nueva y más corta escala de depresión, comparada con la escala de Hamilton o la de Beck para depresión.

- Validez predictiva: Cuando se correlaciona el resultado del instrumento con un desenlace que ocurrirá en el futuro. Por ejemplo, una escala sobre *locus* control externo (personas que tienen a alguien externo que define por ellos) y presentar como desenlace colon irritable en el futuro.

Validez de criterio concurrente

En caso de que exista una medida externa, se pueden correlacionar las diferencias de los puntajes del instrumento evaluado y de la medida externa (por lo general, validada), mediante coeficientes de correlación de Pearson para datos continuos, o de Spearman, para datos categóricos. En caso de que haya una correlación significativa (valores de 0,6-0,9) entre ambos instrumentos, se supone una validez concurrente.

Cabe recordar que si no hay un instrumento validado, se puede utilizar una escala de apreciación clínica subjetiva (por ejemplo, en psiquiatría, la impresión clínica global), o una escala que suministre el nivel de bienestar de 1-100. No obstante, con este tipo de instrumentos es posible que se tenga una más baja correlación de desempeño global.

En caso de no tenerse ningún instrumento, se puede iniciar la validación de dos instrumentos que midan la misma condición en paralelo y hacer la comparación entre ellas, esperando para este tipo de situación que la correlación esté por encima de 0,8. Sin embargo, eso se debe interpretar con precaución.



Las escalas con un muy alto nivel de correlación pueden estar midiendo lo mismo, y podría no haber diferencias entre lo que ambas hacen. Por otra parte, si una escala tiende a tener poca calidad, se esperaría baja correlación si la que se está validando es mejor.

Validez de criterio predictiva

Si el objetivo del instrumento es proponer cómo se desarrollará un evento en el futuro, se habla de validez de criterio predictiva (validez predictiva). En tal caso, la muestra debe ser seguida por un tiempo, hacia el futuro, para determinar la aparición o no del desenlace escogido (por ejemplo, recaída, intento de suicidio, etc.). Se podría plantear que se evalúe si los resultados en las pruebas nacionales de finalización de estudios secundarios predicen las del estudiante durante la carrera universitaria.



En general, para medir la validez de criterio predictiva se puede utilizar la estadística *kappa*, y los niveles recomendados como aceptables son los valores > 0,75 (excelente concordancia); se habla de buena concordancia si los valores están entre 0,4 y 0,75; y se habla de pobre concordancia cuando los valores de *kappa* son < 0,4[29].

Es igualmente digno de anotar que en ambos tipos de validez de criterio los instrumentos deben correlacionarse en un mismo punto del tiempo (validez de criterio de corte transversal).

Validez de constructo

Cuando no existe una medida o patrón de oro, se explora este tipo de validez, la cual procura corroborar modelos, teorías o hipótesis acerca de fenómenos no observables o medibles fácilmente, y que requieren un constructo para explicar las relaciones entre los factores estudiados. Este es el caso de variables como la inteligencia, la ansiedad, el papel de enfermo, etc.

En esta validación se evalúa si la escala discrimina diferentes grupos de sujetos, de una forma que sea coherente con la teoría o el modelo subyacentes. Para ello se aplica el instrumento a grupos con la mayor divergencia posible en la dimensión teórica estudiada, y se verifica si proveen resultados diferentes y congruentes con la teoría.

EJEMPLO

Si un instrumento mide la inclinación a la delincuencia, se aplica en un grupo de personas encarceladas y se compara con los datos obtenidos en un grupo de religiosos.

Para esta validación, el cálculo del tamaño de muestra se hace con los procedimientos regulares (véase el **capítulo 6** "Cálculo del tamaño de muestra"), con una precaución especial para evitar el error tipo II (se recomienda aceptar valores < 10%). Si después de aplicar las pruebas

estadísticas pertinentes aún existe diferencia entre ambos grupos, la validez de constructo estaría confirmada; en caso contrario, el instrumento deberá ser revisado o abandonado.

En caso de que se disponga de otro instrumento o medida que evalúe de alguna forma el constructo teórico estudiado (medida externa), se puede evaluar la correlación que haya entre los dos instrumentos. En caso de que se observe una correlación significativa (> 0,5 o 0,6), se confirma la validez de constructo del instrumento.

En el caso de instrumentos evaluativos, la validez de constructo se evalúa de acuerdo con el cambio en el curso del tiempo (evaluación longitudinal), suponiendo que algún tipo de intervención conlleva, teóricamente, un cambio; esto es, si los cambios, por ejemplo, del estado clínico, se ven reflejados en los puntajes del instrumento. Para ello se deben hacer dos mediciones antes y después del tratamiento o el manejo instaurados, y se comparan los dos puntajes del instrumento mediante pruebas pareadas (Wilcoxon ***signed-rank*, T de Student),** de acuerdo con su pertinencia.

EVALUACIÓN DE LA CONFIABILIDAD

Conocida también como ***reproducibilidad.*** Consiste en la capacidad de la escala para proveer los mismos resultados en mediciones sucesivas, con las mismas características; dicho de otra forma.



Si el instrumento provee resultados similares bajo diferentes condiciones, algunas de las cuales dependen del mismo instrumento, del tiempo o momento de aplicación, y de quien o quienes lo aplican.

Por ejemplo, si a un paciente estable y sin tratamiento se le aplica una escala de dolor en dos ocasiones diferentes, pero dentro de un periodo corto, los resultados deben ser similares en ambas mediciones. Hay tres tipos de confiabilidad, que se relacionan con:

- ***El instrumento:*** **Homogeneidad o consistencia interna.**
- ***El tiempo de aplicación:*** **Test-retest.**
- ***El sujeto:*** **Intraevaluador e interevaluador.**
- Cada una de ellas debe ser evaluada en el proceso de validación. Pasaremos, pues, a su descripción.

Confiabilidad relacionada con el instrumento

Toda esta confiabilidad tiene lo que se denomina ***homogeneidad*, o consistencia interna; ello** se refiere a qué tan coherente o "aglutinada" se encuentra la escala, teniendo en cuenta que los ítems deben tener cierta coherencia entre ellos, y estos, a su vez, con los factores o los dominios que representan, y, por último, los ítems con la escala en general. Si una escala es consistente internamente u homogénea, deberá tener cierta "aglutinación" o coherencia. La consistencia interna será alta si el constructo o la condición medidos son homogéneos y si los ítems miden dicho constructo o condición. En caso contrario, será calificada como ***baja.***

Las medidas de consistencia interna más a menudo utilizadas son el alfa de Cronbach y el coeficiente **KR-20 (Fórmula 20 de Kuder-**Richardson [**KR-20]). La primera se utiliza** para los ítems que pueden responderse en más de dos formas; entre tanto, la segunda **(KR-20) se usa para los ítems con respuesta** dicotómica. En general, se acepta que los valores (índices) entre 0,7 y 0,9 (70 %-90 %) son considerados buenos; no obstante, para muchos fines se pueden aceptar puntajes > 0,6. Valores más bajos hacen pensar en una pobre homogeneidad de la escala, y estar, por ello, midiendo diferentes fenómenos y que cambia ante diferentes circunstancias de aplicación. Ahora bien, una consistencia muy alta (> 0,9) podría reflejar mucha homogeneidad, y con ella, posiblemente, ítems que podrían ser redundantes[30,31].

Por lo anterior, la consistencia interna será alta si el constructo medido es homogéneo y si los ítems miden dicho constructo. En caso contrario, será baja. Finalmente, es importante recordar que la consistencia interna (en especial, el alfa de Cronbach) se puede incrementar artificialmente si el número de ítems es grande.

Confiabilidad relacionada con el tiempo de aplicación

Este tipo de confiabilidad, llamada habitualmente ***test-retest*, explora la capacidad del**

instrumento para dar el mismo resultado en una nueva aplicación después de un lapso pequeño, cuando la condición que se mide no ha cambiado. Para ello, se aplica el mismo instrumento en dos ocasiones, con un intervalo de entre 7 y 30 días (según la condición y el instrumento, puede requerirse un lapso mayor), y se utiliza, para la segunda aplicación, una muestra representativa de la primera (con un mínimo de 30 personas). Posteriormente, se hace la comparación por medio de alguna o algunas de estas medidas:

a. **Coeficiente de correlación de Pearson para** datos continuos.
b. **Coeficiente de correlación por rangos de** Spearman para datos categóricos.
c. **Coeficiente de correlación intraclase (CCI).**
d. **Coeficiente de Lin (coeficiente de correlación-concordancia de Lin).**

Respecto al Pearson, este evalúa de manera lineal los puntajes en diferentes momentos. Asume toda la diferencia encontrada en términos de los sujetos medidos, lo cual lleva a que esta prueba sea poco utilizada, por cuanto no incorpora otras fuentes de variabilidad.

En cuanto al coeficiente de correlación intraclase (**CCI), este se puede calcular usando** un análisis de varianza de mediciones repetidas (**ANOVA de medidas repetidas), e incorpora en** su medición mayores fuentes de variación, como la variabilidad entre sujetos medidos, la variabilidad intrasujeto medido (características de la persona) y el error de medición. El resultado de esta prueba se interpreta como el porcentaje de variabilidad atribuible solo a la variabilidad entre los sujetos medidos. De esta manera, un valor de 0,7, significa que el 70 % se explicaría por cambios entre los sujetos medidos. Se espera que en este tipo de medición se den valores > 0,8 con el fin de proveer una adecuado test-retest.

El coeficiente de correlación-concordancia de Lin, por su parte, es una medición complementaria al CCI, **en esta se mide la concordancia entre dos mediciones**, y, a la vez, la variación en torno a una línea llamada ***dispersograma*** (línea de 45°).

Esta medición provee datos que complementan la anterior.

Confiabilidad relacionada con el sujeto

Este tipo de confiabilidad puede ser intraevaluador e interevaluador. En el primer caso, se refiere a cómo un mismo evaluador (observador) en diferentes ocasiones hace mediciones similares sobre un mismo sujeto que no ha tenido cambios en su condición. Mientras, la confiabilidad interobservador (interevaluador) se refiere a cómo dos o más evaluadores (observadores), con entrenamiento uniforme, proveen resultados similares sobre los mismos sujetos medidos en condiciones similares. En ambos casos se procura medir la variación introducida por los observadores.

Para dichos estudios se establecen previamente el número de observadores, el tiempo en el que se harán las mediciones y el nivel de significancia necesario con el fin de determinar la muestra requerida. La confiabilidad interevaluador se lleva a cabo mediante una evaluación independiente de, por lo menos, quince personas, efectuada por dos o más evaluadores. Los resultados se evalúan con una prueba de concordancia (***kappa*** para datos discretos, o coeficiente de correlación intraclase para datos continuos).

Con el fin de medir la variabilidad intraobservador (intraevaluador), se utilizan pruebas para muestras no independientes, de acuerdo con el tipo de variable (por ejemplo, prueba *T* pareada de Student, prueba Wilcoxon, prueba de Friedman o coeficiente de concordancia de Kendall), y se puede evaluar la varianza intraindividual calculando la desviación estándar de los puntajes de cada sujeto y el promedio de estas entre los sujetos. La magnitud de esta última medida puede suministrar información acerca de la sensibilidad al cambio de la escala.

EVALUACIÓN DE LA SENSIBILIDAD AL CAMBIO, O *CAPACIDAD DE RESPUESTA*

Se refiere a la propiedad de la escala para detectar las variaciones o los cambios que se presentan en el fenómeno medido. Esta característica es fundamental para evaluar estudios de intervención y hacer seguimientos sobre pequeños cambios en los desenlaces. Así, por ejemplo, una escala para medir dolor debe disminuir el

puntaje cuando disminuye la sensación (por un medicamento, por ejemplo, o por mejoría del proceso), y también debe detectar el aumento del dolor, con el fin de detectar cambios que permitan evaluar la intervención (por ejemplo, analgésicos, fisioterapia, etc.).

En general, los instrumentos evaluativos requieren una buena sensibilidad al cambio; es decir, que aun pequeñas variaciones clínicas se vean reflejadas en los puntajes. Para evaluar la sensibilidad al cambio, se requiere aplicar la misma escala de manera repetida en el mismo sujeto; por ejemplo, antes y después del tratamiento.

Es importante considerar que este cambio puede depender de la población de estudio y del tipo de tratamiento. En el primer caso, a medida que la población sea más heterogénea, ello, probablemente, conducirá a puntajes mucho más variables, y en el segundo, cuando se tienen tratamientos que son muy efectivos en algunos pacientes y poco efectivos en otros, se espera encontrar un mayor cambio, comparado con un tratamiento que siempre resulta poco efectivo para el total de los pacientes[32].

Dado lo anterior, algunos autores[4,32] han propuesto el uso de diferentes índices; sin embargo, vamos a mencionar solo los relacionados con los siguientes dos escenarios (que se basan en un grupo único):

1. ***El cambio en los pacientes es homogéneo:*** En tal escenario, el índice es calculado como el cambio promedio dividido por la variabilidad de este cambio. Los estadísticos más utilizados para su cálculo son la prueba ***T*** pareada y el análisis de varianza para medidas repetidas. Adicionalmente, el mencionado índice sirve para calcular el tamaño de muestra necesario para el estudio de instrumentos evaluativos.
2. ***El cambio en los pacientes es heterogéneo:*** En dicho escenario, el índice es calculado con ayuda de un estándar externo, en el que se evalúa el cambio de este y se compara con el cambio de la medida de interés, a través de un coeficiente de correlación de Pearson.

Acerca de la forma de hacer estas pruebas, Guyatt et al.[33,34] proponen que para aumentar la sensibilidad al cambio se le deben mostrar al sujeto sus respuestas previas (anterior aplicación) en el instrumento evaluado. El tiempo que debe ser empleado dependerá de la temática y de las hipótesis que se estén evaluando.

NOTAS FINALES

Finalmente, se obtiene una escala adaptada y validada, que podrá ser utilizada con tranquilidad para medir adecuadamente la condición de interés y los desenlaces. No obstante, a veces es necesario tener un umbral para la calificación, con el fin de hacer positiva o negativa una escala o un síndrome; en tal caso, el instrumento, en caso de serlo, deberá utilizarse como una prueba diagnóstica, y debe seguir los métodos que se disponen para lograr un umbral de esta (véase el **capítulo 10** "Uso de pruebas diagnósticas en medicina clínica").

No sería adecuado culminar el presente capítulo sin decir que este proceso, fundamental para la investigación, requiere un entrenamiento específico en el área, que ayude a contar con los instrumentos adecuados, que sirvan para proveer resultados válidos en los estudios. Si bien no siempre es necesario desarrollar el proceso completo, dado el caso debe haber razones sólidas y explícitas para no llevarlo a cabo.

Referencias

1. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38:27-31.
2. American Educational Research Association, National Council on Measurement in Education, American Psychological Association (APA). Standards for educational and psychological testing. Washington, DC: American Psychological Association; 1985.
3. Nunnally JC. Psychometric theory [internet]. New York: McGraw-Hill; 1978. [citado 2014 jul. 22]. Disponible en: http://books.google.co.in/books?id=WE59AAAAMAAJ
4. DeVellis RF. Scale development: theory and applications [internet]. Newbury Park: Sage Publications; 1991. [citado 2014 jul. 22]. Disponible en: http://books.google.com.co/books?id=hhhHAAAAMAAJ
5. Phillips MR, Yangping Z. Methodological fra-

mework for the evaluation and revision of translated instruments. Clinical Applied Health Social Science, Module 7. Callaghan, Australia: University of Newcastle; 1995.
6. Phillip MR. Evaluation and revision of translated western scales. Rating Scales for Mental Health. 1993. p. 343-51. [Unpublished translation].
7. Spielberger CD, Sharma S. Cross-cultural measurement of anxiety. En: Phillips MR, Yangping Z. Methodological framework for the evaluation and revision of translated instruments. Callaghan, Australia: University of Newcastle; 1995.
8. Sánchez R, Gómez C. Conceptos básicos sobre validación de escalas. Rev Col Psiquiatr. 1998;2:121-30.
9. Cronbach LJ. Essentials of psychological testing [internet]. Harper & Row Limited; 1984. [citado 2014 jul. 22]. Disponible en: http://books.google.com.co/books?id=PvxGAAAAMAAJ
10. Allen MJ, Yen WM. Introduction to measurement theory [internet]. Brooks/Cole Pub.; 1979. Disposnible en: http://books.google.com.co/books?id=cgElAQAAIAAJ
11. DeVellis RF. Scale development: Theory and applications. Applied Social Research Methods Series. Thousand Oaks, CA: Sage Publications; 2003.
12. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Nueva York: Oxford University Press; 2008.
13. Sánchez R, Echeverry J. Validación de escalas de medición en salud. Rev Salud Pública. 2004;6:302-18.
14. Holden RR, Fekken GC. Structured psychopathological test item characteristics and validity. Psychol Assess. 1990:35-40.
15. Huskisson EC. Measurement of pain. The Lancet. 1974;304:1127-31.
16. Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Socioecon Plann Sci. 1976;10:129-36.
17. Norman GR, Streiner DL. PDQ statistics [internet]. PMPH-USA; 2003 [citado 2014 jul. 22]. Disponible en: http://books.google.com/books?hl=es&lr=&id=huoPAHPkxVYC&oi=fnd&pg=PR6&dq=PDQ+Statistics&ots=Nf6PJJAvK9&sig=BAfMegOk6hoytR8M23LkR5nI19M
18. Fidell LS, Tabachnick BG. Using multivariate statistics. New York: Harper Row; 2006.
19. Sepúlveda MAR. Introducción al análisis factorial exploratorio. Rev Colomb Psiquiatr. 2012;41:197-207.
20. Pett MA, Lackey NR, Sullivan JJ. Making sense of factor analysis: The use of factor analysis for instrument development in health care research [internet]. Sage; 2003. [citado 2014 jul. 22]. Disponible en: http://books.google.com.co/books?id=5Jyaa2LQWbQC
21. Thompson B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association; 2004.
22. Hair JF, Tatham RL, Anderson RE, et al. Multivariate data analysis. Nueva Jersey: Pearson Prentice Hall; 2006.
23. Cabrera-Nguyen P. Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. J Soc Soc Work Res. 2010;1:99-103.
24. Ledesma RD, Valero-Mora P. Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out parallel analysis. Pr Assess Res Eval. 2007;12:1-11.
25. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pr Assess Res Eval. 2005;10:173-8.
26. Comrey AL, Lee HB. A first course in factor analysis [internet]. Psychology Press; 2013 [citado 2014 jul. 22]. Disponible en: http://books.google.com/books?hl=es&lr=&id=RSsVAgAAQBAJ&oi=fnd&pg=PP1&dq=A+first+course+in+factor+analysis&ots=I3F8fqmbAe&sig=B-xcjcbjwUGszm5ko5FPV6uh6YQ
27. Worthington RL, Whittaker TA. Scale development research a content analysis and recommendations for best practices. Couns Psychol. 2006;34:806-38.
28. Henson RK, Roberts JK. Use of exploratory factor analysis in published research common errors and some comment on improved practice. Educ Psychol Meas. 2006;66:393-416.
29. Glick H, Williams S. Readings for the health economic course. Philadelphia: University of Pennsylvania; 1994.
30. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53-5.
31. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol. 1993;78:98.
32. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. Heal Qual Life Outcomes. 2005;3:23.
33. Guyatt GH, Berman LB, Townsend M, et al. Should study subjects see their previous responses? J Chronic Dis. 1985;38:1003-7.
34. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40:171-8.

CAPÍTULO

9

Medidas de frecuencia, de asociación y de impacto

Juan Manuel Lozano León

INTRODUCCIÓN

Un elemento común en casi todos los estudios clínicos y epidemiológicos es la necesidad de expresar la frecuencia de un evento de interés. Más aún, en muchas oportunidades se requiere comparar la frecuencia observada de un fenómeno, en varios grupos, para establecer si hay cambios respecto a la presencia de otras condiciones —la exposición a un posible factor causal o a una intervención preventiva o terapéutica—. Este capítulo discute las medidas para describir la frecuencia de un fenómeno en una población, así como las medidas para establecer la asociación entre ciertas variables y la frecuencia del evento.

GENERALIDADES: RAZONES Y PROPORCIONES

La forma más común de expresar la frecuencia de un evento es como una razón: una cifra (numerador) dividida por otra (denominador).

La naturaleza de cada una de estas dos cifras cambia según la medida de frecuencia o de asociación específica, como ocurre con las proporciones, los porcentajes, las tasas, etc. Es importante recordar que en algunos casos, como en las proporciones o en los porcentajes, el numerador está incluido dentro de la población del denominador, mientras que no ocurre lo mismo en otras razones[1].

EJEMPLO

Se tiene una población de 200 personas, en la cual hay 40 mujeres. La frecuencia de mujeres puede expresarse como una proporción (40/200 = 0,2), como un porcentaje (40/200 × 100 % = 20 %) o como la razón mujeres/hombres (40/160 = 0,25). Es claro que si bien en las dos primeras expresiones las 40 mujeres forman parte de los 200 sujetos del denominador, no ocurre así con la razón, cuyo denominador solo está compuesto por la población de hombres.

Las unidades de medida, o ***dimensiones***, de una razón corresponden a la adición o la cancelación algebraica de las dimensiones del numerador y del denominador. A menudo, di-

cha cancelación hace que la razón resultante no tenga dimensiones, como también ocurre con las proporciones. En otras oportunidades, las razones se expresan como porcentajes, o, en otros casos, con unidades más complejas.

EJEMPLO

Los resultados de las pruebas de función pulmonar se dan en varias razones: el volumen espiratorio forzado durante el primer segundo (VEF_1) corresponde a la cantidad de aire exhalada durante el primer segundo de la espiración forzada, y sus unidades son litros. Si se quiere obtener la proporción del VEF_1 respecto a la capacidad vital forzada (VEF_1/CVF), las unidades en cada uno de ellos (litros) se cancelan para dar una proporción sin dimensiones, que puede multiplicarse por 100 para expresarla como porcentaje. Por último, el flujo espiratorio a la mitad de la CVF ($FEF_{0,5}$) tiene como unidades litros/segundo.

En las páginas que siguen se describen con más detalle algunas de las proporciones o razones utilizadas comúnmente en investigación clínica o epidemiológica, así como su interpretación.

MEDIDAS DE FRECUENCIA

Las dos medidas de frecuencia más utilizadas en la investigación clínica y epidemiológica son la ***prevalencia*** y la ***incidencia***. Si bien en el lenguaje común a veces se los emplea como sinónimos, estos dos términos tienen significados por completo diferentes.

Prevalencia

El concepto de prevalencia se refiere al número de veces que se encuentra una condición en una población, en un momento determinado[2-4].

Las prevalencias, que son proporciones y no tienen dimensiones, se expresan usualmente como porcentajes, aunque para eventos menos comunes se pueden emplear denominadores más grandes, como el número de casos observados en 1000, 10 000 o 100 000 personas. Los estudios para establecer la prevalencia se limitan a determinar la presencia o la ausencia del evento de interés (que usualmente es algo indeseable como estar enfermo, ser seropositivo para alguna infección, fumar, etc.) en cada miembro de una población, calculando el numerador (total de enfermos, de seropositivos, de fumadores, etc.) y el denominador (total de personas que componen la población). Tales estudios se llaman estudios de corte transversal, encuestas o, simplemente, estudios de prevalencia.

Como ya se describió en el **capítulo 14** "Estudios de corte transversal", las principales dificultades metodológicas de esas investigaciones radican en la selección de las estrategias para identificar a los enfermos y para evaluar al total de la población, así como en la definición del periodo durante el cual se va a llevar a cabo la medida.

EJEMPLO

Se hizo un estudio para determinar la prevalencia de la disfunción eréctil en varones de 40 o más años en Colombia, Ecuador y Venezuela. Luego de elegir al azar una muestra de 1963 hombres, en 10 ciudades de los 3 países, se procedió a realizar una entrevista, con un instrumento validado para detectar disfunción eréctil. Dicho estudio mostró que la prevalencia, ajustada por edad, es del 33,6 %, del 16,2 % y del 3,6 %, respectivamente, para trastorno mínimo, moderado y completo[5].

Es importante que el investigador defina el total de la población, pues sin esa cifra no se puede establecer el denominador de la prevalencia. Esta población cambia de acuerdo con el objetivo del estudio, y puede ser el total de sujetos hospitalizados en un servicio clínico, el de los pacientes que consultan a una institución, el de donantes en un banco de sangre o el de la población de una determinada región geográfica (una ciudad, un país, etc.), para mencionar unos pocos ejemplos. De la misma forma, en algunas circunstancias el investigador deberá definir las características de la población en términos de edad, sexo, grupo racial, etc., según como convenga para los objetivos del estudio. Cuando no se puede tener acceso a la totalidad de la población de interés, se debe emplear un muestreo aleatorio, que permite generalizar los resultados a la población fuente de donde se tomó la muestra.

Por último, según la extensión del periodo durante el cual se hagan las mediciones, la prevalencia puede adoptar diferentes formas. En algunas circunstancias se puede medir la prevalencia en un momento muy específico en el tiempo. Por ejemplo, un investigador podría establecer la prevalencia de la neumonía nosocomial en los pacientes que se encuentran en la Unidad de Cuidado Intensivo (UCI) de un hospital en una hora y una fecha determinadas. Para ello, basta con contar el número de casos que cumplen los criterios para el diagnóstico de neumonía nosocomial y dividirlos por el total de pacientes hospitalizados en la unidad en ese mismo instante. Esta es una ***prevalencia de punto***[2-4]. En otras oportunidades, la población que se va a evaluar es muy grande y no se puede hacer el estudio en poco tiempo; por eso, cuando la evaluación toma varios días, semanas o, incluso, meses, se habla de *prevalencia de periodo*[2-4]. Finalmente, se pueden encontrar ***prevalencias anuales***, donde se reporta el total de individuos que tienen la enfermedad en cualquier momento a lo largo de un año, y se incluyen tanto los casos que empezaron antes y que aún están enfermos al comenzar la medición como los de quienes se enfermaron durante dicho año. En estos tres casos, se reconoce que la medida sigue siendo prevalencia, pues el estado de cada individuo (caso o no caso), en la población o en la muestra seleccionada es documentado solo una vez durante el periodo de interés (de punto —instante preciso—, de periodo —lapso— o anual), a diferencia de la medición de incidencia, donde en el estudio se documenta si el estado de cada individuo cambia a lo largo de un periodo, tal como se explica en la siguiente sección.

Incidencia

La incidencia corresponde al número de personas que presentan un fenómeno (también, usualmente, algo indeseable, como enfermar, recaer o morir) durante un periodo determinado y en una población inicialmente libre del proceso y que se encuentra en riesgo de presentarlo.

A diferencia de la prevalencia, cuyo numerador es el número de casos presentes en la población en un momento determinado, la incidencia se concentra en el número de eventos ***nuevos***[2-4]. En consecuencia, su numerador es el total de casos nuevos que aparecen en la población, y se descartan los que ya existían al inicio del periodo de observación.

Esta diferencia hace que los estudios para determinar la incidencia sean más complejos que los de prevalencia, pues implican el seguimiento periódico a un grupo de individuos, inicialmente libres de la enfermedad, para identificar los nuevos eventos a medida que ocurren. Dichos estudios se conocen como ***estudios de cohorte***[2], y se describen en el **capítulo 17** "Estudios de cohorte analítica". Durante su conducción se deben observar los mismos principios metodológicos descritos para la prevalencia respecto a la identificación de los casos nuevos.

Si la población que constituye el denominador permanece constante a lo largo del tiempo de observación, la acumulación de nuevos eventos dará lugar a una ***incidencia acumulativa***[2-4], que es una proporción, y, por tanto, también se la denomina ***proporción de incidencia***. Sin embargo, no es raro que la población cambie a través del seguimiento, debido a la llegada de nuevos individuos susceptibles o a la pérdida de otros por migración o eventos competitivos. De esta manera, la duración del seguimiento puede ser diferente para cada miembro de la población. Para permitir que cada individuo contribuya durante todo el tiempo de seguimiento la medida de tiempo en el denominador no son los individuos en riesgo, sino la combinación de cada sujeto con su periodo de seguimiento, lo que da lugar a una expresión llamada ***persona-tiempo***.

EJEMPLO

Un individuo seguido durante diez años aporta diez años-persona al denominador, mientras que otro seguido solo por dos años aporta dos años-persona. Si estos fueran los dos únicos miembros de la población, el total de tiempo-persona sería de doce años-persona.

Esta incidencia, que se expresa, entonces, como el número de eventos nuevos (numerador) observados en el total de años-persona de seguimiento (denominador), se llama ***densidad de incidencia***[2-4], y matemáticamente corresponde a un tipo especial de razón o cociente, denominado *tasa*. Por tal motivo, muchos autores se refieren a este cociente como ***tasa de incidencia***, y, en términos prácticos, tasa y densidad de incidencia se pueden considerar sinónimos.

EJEMPLO

Un estudio conducido con todos los ingresos a ocho unidades de neonatos en Bogotá y en Cali a lo largo de un periodo de seis meses permitió establecer la incidencia de infección nosocomial. En dicho estudio de cohorte concurrente se hallaron 127 episodios de infección, confirmada mediante cultivo en 80 neonatos, entre un total de 1504 ingresos que se consideraron en riesgo, por permanecer hospitalizados durante al menos 72 horas, y quienes no tenían evidencia de infección al ingreso. La forma más simple de expresar la incidencia sería tomar el número de sujetos o de infecciones, y dividirla por el total de sujetos en riesgo de adquirir la infección a lo largo de las primeras 72 horas. Esto representaría una incidencia de 5,3 neonatos infectados o de 8,4 infecciones nosocomiales por cada 100 ingresos de más de 72 horas durante el periodo de 6 meses. Sin embargo, dado que la duración de la hospitalización (y, por tanto, del riesgo) es muy variable en los neonatos, el estudio también midió la densidad de incidencia, y la encontró en 4,8 episodios de infección nosocomial por cada 1000 neonatos-día[6].

Usos de los estudios de prevalencia y de incidencia

La información suministrada por los estudios descritos se emplea en los campos clínico y epidemiológico para tres propósitos fundamentales: 1) predecir la frecuencia futura de eventos en individuos o poblaciones, 2) estimar la probabilidad de ocurrencia de un evento en individuos y 3) hacer comparaciones de los resultados con otros estudios similares.

Al describir la frecuencia de aparición de un evento en una población libre de este al inicio del seguimiento, la incidencia provee una muy buena idea sobre qué tan a menudo dicho fenómeno aparecerá a lo largo del tiempo en sujetos similares. En tal sentido, la incidencia es una excelente medida del riesgo que tienen los individuos de una población de presentar el evento en el futuro[7].

La prevalencia, por otra parte, se limita a decir cuál es la probabilidad de que un sujeto, seleccionado al azar de dicha población, tenga el evento o la enfermedad de interés ***en ese momento*, no en el futuro. Esto último, sin embargo**, también es de gran utilidad; en especial, durante el proceso de diagnóstico, pues indica cuán factible es que un individuo presente dicha condición en ese instante (probabilidad previa al examen o al inicio del tratamiento; en general, "probabilidad ***a priori***"), lo que ayuda a tomar decisiones diagnósticas y terapéuticas[8].

Si bien la prevalencia y la incidencia ofrecen información valiosa, como la descrita en el párrafo previo, su mayor aporte reside en su empleo para hacer comparaciones. La prevalencia puede ayudar a planificar los servicios de salud de una ciudad o una región, pues señala cuáles son las enfermedades que más a menudo afectan a una población. Para ello, basta con medir la prevalencia de diversas dolencias en la misma población y compararlas entre sí: las más elevadas representan las entidades que más aquejan a la población, y, por ende, son las que deberían, al menos en términos de su frecuencia, recibir mayor atención. La comparación de la incidencia de un evento en diversas poblaciones, por su parte, permite explorar el efecto que la exposición a ciertas condiciones tiene sobre el riesgo futuro de desarrollar dicho desenlace. Este aspecto y, en especial, las medidas que se emplean para ello se describen más detalladamente a continuación.

MEDIDAS DE ASOCIACIÓN

Uno de los principales objetivos en muchos estudios clínicos y epidemiológicos es determinar si la exposición a algún factor (un posible agente causal, una intervención preventiva, un tratamiento) modifica la frecuencia de aparición de uno o de varios desenlaces de interés. En lenguaje algo más técnico, se pretende establecer si hay alguna asociación entre la exposición al factor de interés y el riesgo posterior de presentar un desenlace clínico indeseable. Esto implica comparar la frecuencia de dicho evento en grupos de sujetos expuestos y no expuestos al agente o a la intervención de interés. Otros capítulos de este libro describen y explican los detalles metodológicos de tales estudios. En los párrafos que siguen se describirán las medidas de resumen más utilizadas para establecer si, en efecto, dichas asociaciones se hallan presentes.

Como se dijo en las secciones precedentes, la mejor medida del riesgo de desarrollar un evento es la incidencia; o sea, la frecuencia con la cual se presenta un desenlace durante un periodo determinado en individuos que estaban libres de él al inicio de la observación.

La comparación de las incidencias observadas en individuos expuestos y no expuestos permitirá establecer cuál es el aumento o la disminución en el riesgo del desenlace según la exposición.

EJEMPLO

Para ilustrar las diferentes medidas que se pueden emplear, se utilizarán los datos de un estudio que evaluó el impacto del sulfato de magnesio, en comparación con placebo, en gestantes con preeclampsia, llamado el *Experimento Magpie*[9]. Un grupo internacional (que incluyó a investigadores de 4 centros en Colombia) condujo un experimento clínico aleatorizado, en el que gestantes con preeclampsia fueron asignadas al azar para recibir sulfato de magnesio (5068 mujeres) o placebo (5068 mujeres). Fueron seguidas 5055 participantes de cada grupo, con el fin de establecer la frecuencia de aparición de varios eventos de interés. El principal fue la frecuencia de eclampsia; es decir, la aparición de convulsiones. Otros desenlaces incluyeron la muerte, por cualquier causa, de la gestante o de su hijo, y la admisión de los neonatos a cuidado intensivo neonatal, entre varios más. Durante el estudio, los investigadores encontraron un total de 40 casos de eclampsia en las mujeres tratadas con sulfato de magnesio, y de 96 casos, en el grupo tratado con placebo.

La **figura 9.1** presenta la forma de analizar los datos de experimentos como este para determinar si hay asociación entre el tratamiento recibido y el riesgo del desenlace. Se parte de sujetos elegibles, en riesgo de presentar el evento, quienes son distribuidos al azar para que reciban una de las dos intervenciones. La línea horizontal (marcada con la letra ***X***) señala la división de la población según su exposición: mientras el grupo ubicado arriba de esta línea se expuso a la intervención de interés (sulfato de magnesio en el caso del experimento Magpie), el ubicado por debajo de ella corresponde a los sujetos del grupo control (que recibió placebo).

Por otra parte, la línea vertical ***Y*** divide estos dos grupos según la aparición del desenlace: los casos ubicados a la izquierda son quienes presentan el desenlace (en Magpie, las mujeres que presentaron eclampsia), en tanto que los de la derecha son quienes no lo presentaron. La división de la población del estudio por esas líneas, que señalan la distribución de la población según la exposición recibida y la aparición ulterior del desenlace, permite resumir los datos en una tabla de dos por dos celdas. Cada celda (identificada con las letras: ***a***, ***b***, ***c*** y ***d***) contiene el total de sujetos que presentaron y no presentaron el desenlace en cada grupo de tratamiento.

La **tabla 9.1** muestra los datos para el experimento Magpie en esta tabla de 2 × 2. A partir de ella, se puede calcular el riesgo (en forma de proporciones) de eclampsia en las gestantes tratadas con sulfato de magnesio o con placebo, que corresponden a 0,008 y 0,019, respectivamente; o, si se quiere, al 0,8 % y al 1,9 %, respectivamente **(tabla 9.2)**. Debe tenerse la precaución de mantener siempre el orden propuesto de las filas (***exposición al factor de interés*** en la fila superior, y ***no exposición*** en la fila inferior) y de las columnas (***desenlace indeseable de interés*** en la izquierda, y ***no desenlace*** en la derecha), ya que de otro modo las fórmulas presentadas en la **tabla 9.2** serían incorrectas.

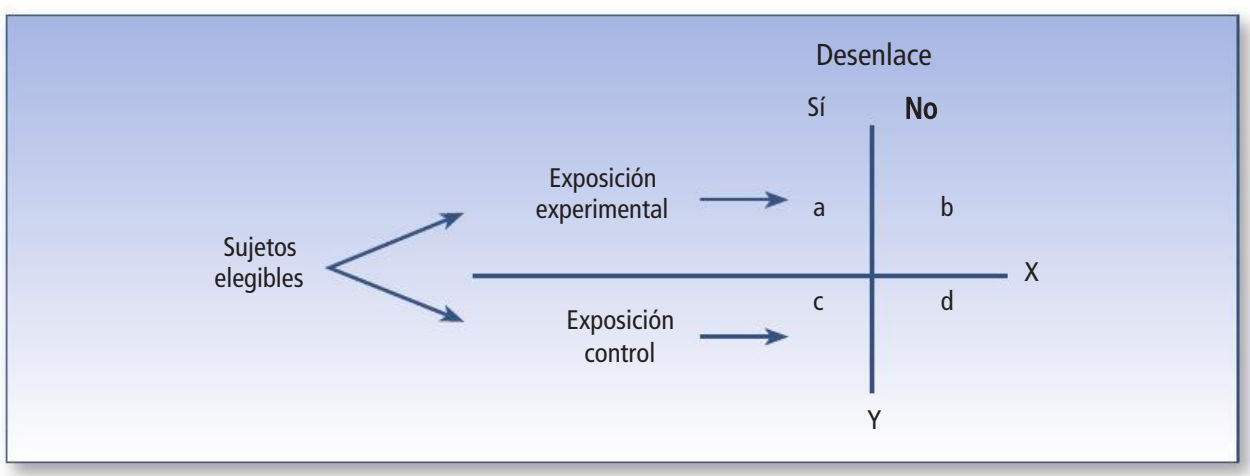


FIGURA 9.1 Evaluación de riesgos en un estudio experimental.

Tabla 9.1 Tabla de 2 × 2 celdas con los resultados del experimento Magpie

	Eclampsia		Total
	Sí	No	
Sulfato de magnesio	40 **a**	5015 b	**5055**
Placebo	c 96	**d 4959**	**5055**
Total	136	9974	10110

Fuente: Magpie[9].

Reducción absoluta del riesgo

¿Cómo se pueden comparar el riesgo hallado en el grupo tratado con sulfato de magnesio y el riesgo en el grupo tratado con placebo, para determinar si son diferentes? Y en caso de que lo sean, ¿cómo se pueden medir la magnitud y la dirección de la dicha diferencia? Lo primero que puede hacerse es, simplemente, establecer la diferencia en la frecuencia de eclampsia observada entre los grupos. Para ello, basta con tomar la frecuencia del evento en el grupo no expuesto (el control que recibió placebo) y restar de ella la frecuencia observada en el grupo expuesto (que recibió sulfato de magnesio). Eso corresponde a la reducción absoluta del riesgo (**RAR**) **observada** por efecto de la exposición[1][7,10].

[1] **Esta medida es similar al riesgo atribuible de es**tudios de causalidad, en los cuales se espera que la exposición aumente el riesgo del evento indeseable, en vez de reducirlo. La diferencia estriba en que el orden de las cifras al establecer la diferencia se invierte: de la frecuencia en los pacientes expuestos se sustrae la de los no expuestos. La interpretación es la misma: muestra la diferencia en los riesgos que se observa como consecuencia de la exposición, y, en términos generales, ambos índices (reducción absoluta de riesgo y riesgo atribuible) son "diferencias de riesgo", que pertenecen a la familia de las llamadas "fracciones atribuibles", como se discute más adelante.

Tabla 9.2 Fórmulas para calcular medidas de asociación, y sus resultados al aplicarlas al riesgo de eclampsia en el experimento Magpie

Medida	Fórmula	Experimento Magpie
Proporción (riesgo) de eventos Expuestos (pe) No expuestos (pne)	 **pe = a/(a+b)** pne = c/(c+d)	 **40/5055 = 0,008** 96/5055 = 0,019
Reducción absoluta del riesgo (**RAR**)	**p_{ne}–p_e**	0,019 – 0,008 = 0,011
Riesgo relativo (**RR**)	pe/pne = a/(a+b)/c/(c+d)	0,008/0,019 = 0,42
Reducción (cambio) del riesgo relativo (**RRR**)[a]	[(pne–pe)/pne] 100 % = [1**RR**]**100 %**	**[0,011/0,019]** 100 % = 58 % = [1 – 0,42] 100 % = 58 %
Suertes de eventos Suertes en expuestos Suertes en no expuestos	a/b c/d	**40/5015 = 0,008** 96/4959 = 0,019
Razón de suertes, o riesgo relativo indirecto *odds ratio* (**RRI**)	**(a/b)/(c/d)** = ad/bc	**0,008/0,019 = 0,42**
Número necesario para tratar (**NNT**)	**1/RAR**	1/0,011 = 91

[a] **Esta medida corresponde a una familia de mediciones conocidas como fracciones atribuibles, y, específicamente, la fracción de** exceso de casos (4), y que, a su vez, en el caso del RR se puede denominar cambio del efecto relativo, el cual puede ser positivo o negativo y se expresa como porcentaje. Cuando es positivo, se denomina exceso relativo de riesgo, y cuando es negativo es RRR. En general, se computa sustrayendo 1 a la razón de medidas de frecuencia relativa y multiplicando esta diferencia por 100. Por ejemplo, en el caso de un estudio de cohorte de factores de riesgo que muestre que la incidencia acumulada de infarto de miocardio es 4 veces mayor en pacientes fumadores (expuestos) que en no fumadores (no expuestos), la razón entre las dos incidencias, o RR, es 4:1, o, simplemente, 4. ¿En qué magnitud se incrementó el riesgo en los fumadores? En (RR-1) **× 100, o (4-1)** × 100 = 300 %. Multiplicar por 4 el riesgo (la frecuencia esperada) es lo mismo que incrementarlo en el 300 % (el 100 % basal, más el 300 % adicional da al final el 400 %, o 4 veces más). En general, para calcular la RRR se puede usar la misma fórmula, donde se sustrae 1 del RR y se multiplica por 100. Se debe tener en cuenta que el signo negativo de la sustracción indica que se redujo el "exceso de riesgo", o que el cambio relativo del riesgo fue negativo. Por tal motivo, en la literatura metodológica sobre el cálculo de la RRR (particularmente, en experimentos que evalúan la efectividad de intervenciones, y en los cuales se espera que la exposición disminuya el riesgo; es decir, que el RR sea < 1), se acostumbra invertir la fórmula: (1–RR) **× 100; y si la diferencia es positiva, se** interpreta como la magnitud de la RRR, tal como se muestra en esta tabla, en vez del confuso "exceso negativo" de riesgo.

Fuente: Magpie[9].

Como puede apreciarse en la **tabla 9.2** para el experimento Magpie, la diferencia fue de 0,011. Un resultado positivo como este significa que el desenlace fue más común entre los pacientes no expuestos que entre los expuestos. Un resultado negativo (en otras palabras, < 0) señalaría lo contrario: el evento indeseable es más común en los pacientes expuestos que en los no expuestos. Finalmente, un valor de cero indicaría que la incidencia es igual en los dos grupos.

La **RAR, sin embargo, no tiene un significado** clínico muy evidente. Por otra parte, como se verá más adelante, la **RAR es útil para calcular otras** medidas que describen el impacto clínico potencial que se puede esperar de la intervención.

Riesgo relativo

Uno de los indicadores más a menudo empleados para comparar los riesgos es el *riesgo relativo* (RR), que corresponde a la división del riesgo del evento en el grupo expuesto por el de los no expuestos (o sea, estudiar el riesgo en los pacientes expuestos *en relación* con el de los no expuestos)[4,7,10-13].

La cifra para el experimento Magpie corresponde a 0,42 **(tabla 9.2). Un resultado < 1,0,** como este, significa que el riesgo en los pacientes expuestos es inferior al de los no expuestos, o, lo cual es equivalente, que la exposición se asocia a una reducción del riesgo. Lo contrario ocurre si se obtiene una cifra > 1,0. Finalmente, un resultado igual a 1,0 significa que el riesgo es igual en pacientes expuestos y no expuestos, lo cual sugiere que no hay asociación entre la exposición y el riesgo del evento.

Reducción del RR

A partir del **RR** se puede establecer el porcentaje de aumento o de reducción en el riesgo observado como consecuencia de la exposición, tomando como base el riesgo de los pacientes no expuestos. Esta expresión, que se conoce como la ***reducción del riesgo relativo*** **(RRR), arroja un** resultado del 58 % en el experimento Magpie, lo cual significa que las mujeres expuestas al sulfato de magnesio tuvieron un riesgo de eclampsia, luego de recibirlo, que fue el 58 % inferior al observado en los sujetos que recibieron placebo **(tabla 9.2)**[10-13].

Esta última medida tiene un significado clínico más evidente, pues indica en cuánto aumenta o se reduce el riesgo de presentar el evento indeseable como consecuencia de la exposición, al compararlo con el riesgo observado en los pacientes no expuestos. Sin embargo, también tiene limitaciones: se observaría la misma **RRR** si los riesgos de los grupos experimental y de control hubieran sido 10 veces mayores (8 % y 19 %).

El RR indirecto

En los párrafos precedentes se estableció la proporción de eclampsia durante el seguimiento en los dos grupos del experimento Magpie (expuestos al medicamento o al placebo). Por corresponder a proporciones, el total de individuos que fallecieron está, igualmente, incluido en el denominador. Otra forma de expresar la frecuencia de muertes en cada grupo es establecer la razón de eclampsias/no eclampsias, y excluir a las primeras del denominador[1,3,4,14]. Estas razones, que se conocen como ***momios***, *suertes* **u *odds*, corresponderían a 0,008 y a** 0,019 para los grupos de sulfato de magnesio y los de placebo, respectivamente **(tabla 9.2).**

Como puede verse, esta es una forma alternativa de calcular el riesgo de presentar el evento, y que, simplemente, determina cuál es la probabilidad de presentarlo dividida entre la probabilidad de no presentarlo.

La **tabla 9.2 muestra que para el experimento** Magpie los riesgos y las suertes son muy similares. Esto, sencillamente, se debe a que el evento de interés (eclampsia) fue poco común en ambos grupos, si bien menos en los pacientes expuestos que en los no expuestos.

La **figura 9.2 muestra que las suertes tenderán** a ser más parecidas a la proporción correspondiente cuanto más raro sea el evento de interés, pues si el evento es poco común, los denominadores de la proporción y de las suertes serán muy similares. Por ejemplo, 1/100 y 1/(100 – 1) arrojan resultados prácticamente idénticos (0,01 y 0,0101, respectivamente), en tanto que 80/100 y 80/(100 – 80) son muy diferentes (0,8 y 4,0, respectivamente). La **tabla 9.3**

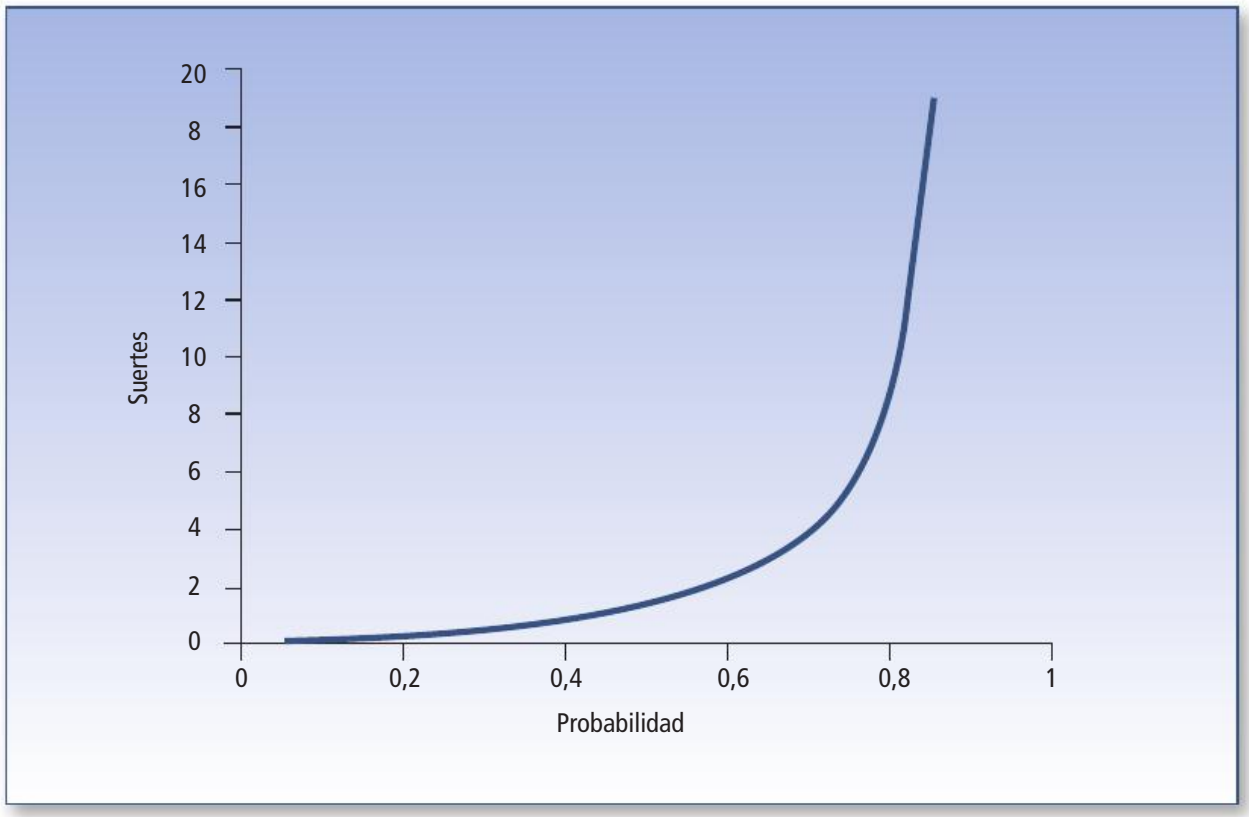


Figura 9.2 Evaluación de riesgos en un estudio experimental.

Tabla 9.3 Fórmulas para calcular las suertes a partir de una proporción y viceversa y sus resultados al aplicarlas al riesgo de eclampsia en el experimento Magpie

Objetivo	Fórmula	Experimento Magpie
Obtener las suertes a partir de una proporción	**suertes** = $p/(1-p)$	$suertes_e$ = 0,008/(1 – 0,008) = 0,008 $suertes_{ne}$ = 0,019/(1 – 0,019) = 0,019
Obtener la proporción a partir de las suertes	p = suertes/(1+suertes)	p_e = 0,008/(1 + 0,089) = 0,008 p_{he} = 0,019/(1 + 0,019) = 0,019

Fuente: Magpie[9].

presenta las fórmulas generales para obtener las suertes a partir de proporciones o para obtener proporciones con base en las suertes.

Así como se puede obtener el riesgo relativo, se puede calcular la razón de suertes (***odds ratio*** en inglés), la cual no es más que las suertes del evento en los pacientes expuestos, divididas por las suertes del evento en los no expuestos **(tabla 9.2)**[1,4,12-15]. Esto corresponde a un riesgo relativo indirecto. Se le llama indirecto porque, estrictamente hablando, no se utilizan las proporciones del evento, que son los mejores indicadores de los riesgos, sino las suertes, que son una aproximación a ellos[2].

La interpretación de este resultado es idéntica a la del **RR: un valor < 1,0 sugiere que la** exposición se asocia a una reducción en el riesgo de presentar el evento, mientras que cifras

[2] **La razón de momios es un cociente de cocientes,** entre las razones de frecuencias absolutas o relativas de evento y no evento entre dos grupos (por ejemplo, pacientes expuestos y no expuestos). Tanto la razón de riesgos (**RR**) como la razón de momios expresan cuantitativamente la magnitud de la asociación entre

iguales a 1,0 o mayores implican que no hay asociación, o bien, que la exposición aumenta el riesgo, respectivamente.

Como puede verse, en estudios experimentales o de cohorte la asociación entre la exposición y el riesgo puede evaluarse calculando tanto el RR como la razón de suertes (o RR indirecto). Sin embargo, en los estudios de casos y controles solo puede obtenerse el RR indirecto.

Ello se debe a que este diseño permite determinar la frecuencia de la exposición en sujetos que ya tienen el desenlace de interés (los casos), comparándola con la observada en individuos que no han presentado el desenlace (los controles). Así, los estudios de casos y controles no miden el riesgo del desenlace, pues para ello es necesario iniciar el estudio con individuos libres del desenlace que son seguidos a lo largo del tiempo, lo cual solo es factible en los estudios de cohorte o en los experimentos aleatorizados.

El **RR**, bien sea obtenido de forma directa o indirecta, es una medida de la fortaleza o de la magnitud de la asociación entre la exposición y el desenlace. Teóricamente, sus resultados varían entre 0,0 y el infinito. Cuanto más alejado esté el resultado de 1,0, mayor será la asociación. Por ejemplo, valores como 0,5 y 2,0 indican una reducción del riesgo a la mitad (al 50 %) o un aumento de este al doble (del 100 %) como consecuencia de la exposición. Resultados iguales a 0,1 y 10 señalarían una reducción a la décima parte (o al 10 %) y un aumento de 10 veces (del 1 000 %), respectivamente. Dicha información es muy valiosa, pero debe ponerse en contexto, teniendo en cuenta la precisión del valor calculado. Este concepto se discute con mayor detalle más adelante en este mismo capítulo.

diferentes exposiciones y un desenlace, pero lo hacen en ***escalas*** diferentes. Solo cuando se cumple la condición de que la frecuencia del desenlace (las cifras en las casillas A y C de las tablas de 2 × 2, como la ilustrada en la **figura 9.1**, sean muy pequeñas, comparadas con B y D (es decir, que el evento de interés sea muy poco común) las magnitudes del **RR** y de la razón de momios resultan numéricamente muy cercanas, y el valor de una de ellas es una aproximación aceptable de la otra. A medida que los valores de A y C aumentan en comparación con B y D (la frecuencia del evento aumenta), el valor numérico de la asociación diverge: mientras que el cociente de riesgos (**RR**) no cambia, el cociente entre los momios aumenta, tal como lo muestra la **figura 9.2**.

Número necesario para tratar

Finalmente, una medida que permite obtener una idea del posible impacto clínico de la intervención en la práctica es el número de sujetos a quienes es necesario tratar para evitar un desenlace desfavorable (NNT)[10-12] **(tabla 9.2)**. El **NNT** corresponde al inverso de la **RAR**.

En el caso de estudios como Magpie, que pretenden evaluar la efectividad de una intervención preventiva o terapéutica, esta razón nos indica cuántos individuos deben recibir la intervención para evitar un desenlace desfavorable. Según los datos del estudio en mención, aproximadamente 91 gestantes con preeclampsia deben recibir el medicamento para prevenir una eclampsia, en tanto que 562 deben recibir tratamiento para evitar una muerte materna.

El tratamiento ideal (que no existe) tendría un **NNT** de 1, lo cual indica que por cada sujeto tratado se evita un desenlace desfavorable, mientras que una intervención que no es superior a otra tendría un **NNT** equivalente al infinito.

En estudios donde la exposición se asocia a un incremento en el riesgo de eventos indeseables, el **NNT** tendría un resultado negativo, que representa el total de personas que deben exponerse al factor de riesgo para observar un evento indeseable. Algunos autores han propuesto emplear en tales circunstancias el término NNH (de la expresión ***number needed to harm*** en inglés).

Las frecuencias de debilidad muscular en las mujeres que recibieron sulfato de magnesio y placebo en el experimento Magpie fueron del 1,4 % y el 0,12 %, respectivamente. Ello representa una **RAR** de – 0,0132, lo que equivale a un NNH de 76. En otras palabras, según los datos del mismo estudio, una de cada 76 mujeres tratadas presentará debilidad muscular. La comparación de este NNH con el **NNT** descrito en el párrafo previo permite examinar el balance entre los beneficios y los riesgos de esta intervención: por cada 1000 mujeres con preeclampsia tratadas con sulfato de magnesio se evitarán 11 casos de eclampsia a expensas de producir debilidad muscular en 13 gestantes.

Debido a que tanto el **NNT** como el **NNH** son medidas absolutas del riesgo de ocurrencia de

eventos, pues son el inverso de la diferencia de riesgos (la diferencia de proporciones de incidencia), su valor depende no solo de la magnitud relativa del efecto de la exposición (del RR), sino también, de la frecuencia basal del evento, que varía entre diferentes poblaciones. Por tanto, es inadecuado comparar NNT entre diferentes estudios cuyas poblaciones de interés difieran en el riesgo basal del evento, en el tiempo de seguimiento o en el espectro y la definición de los desenlaces de interés[15].

EJEMPLO

Se hace un estudio para examinar el efecto de la morfina administrada en una dosis única fija, para controlar adecuadamente el dolor (a las 2 horas del postoperatorio en una cirugía de reemplazo de cadera), por comparación con una dosis fija de un antiinflamatorio no esteroideo (AINE) en 200 pacientes. De los 100 sujetos en AINE, 50 reportan inadecuado control, comparados con solo 10 de los 100 pacientes con morfina. El RR de inadecuado control del dolor es 0,1/0,5 = 0,2. La RAR es 0,5 – 0,1 = 0,4 y el NNT es 1/0,4 = 2,5; o sea, hay que tratar, en promedio, a 2,5 sujetos con una dosis de morfina para tener un adecuado control del dolor en un sujeto adicional, si se compara con tan solo haberles administrado una dosis de AINE.

Supongamos que se hizo el mismo tipo de experimento, pero comparando una dosis única de tramadol con el mismo AINE, a la misma dosis fija, en el mismo tipo de cirugía. Los participantes en tal estudio tienen una peor tolerancia al dolor, reflejada en el hecho de que, a diferencia del estudio anterior (donde 50 de cada 100 sujetos reportaban persistencia del dolor con AINE), en el presente estudio la reportan 80 de cada 100. Con tramadol, solo 16 de cada 100 reportan que persiste el dolor a las 2 horas del postoperatorio. En este caso, el RR es 0,16/0,8 = 0,2; o sea, igual al que se obtuvo al comparar morfina con AINE. La RAR es 0,8 – 0,16 = 0,64; esta es mayor, en términos absolutos, que la obtenida en el primer estudio, y, consecuentemente, el NNT es más pequeño: 1/0,64 = 1,56.

Si se comparan los RR y las RRR de ambos estudios, los datos sugieren que aun cuando la frecuencia basal de intensidad o de tolerancia al dolor sea diferente (con la misma dosis de AINE, las frecuencias de persistencia del dolor en las dos poblaciones fueron diferentes), las magnitudes de los efectos de la morfina y del tramadol son iguales: $RR = 0{,}2$ y $RRR = (1 - 0{,}2) \times 100 = 80\,\%$ de reducción en la incidencia de desenlaces desfavorables.

Si se opta por comparar las RAR y los NNT que de ellas se derivan, se tiene la falsa impresión de que el tramadol es más efectivo: la reducción absoluta de la frecuencia de mal control del dolor es de 64 puntos porcentuales, mientras que con morfina es de 40 puntos porcentuales y el NNT es más bajo: solo se necesita tratar con tramadol a 1,56 pacientes para tener a un sujeto adicional con buen control del dolor, mientras que con morfina es necesario tratar casi el doble: 2,5 pacientes, en promedio, para un éxito adicional.

Para hacer una comparación justa usando el NNT, los dos medicamentos (morfina y tramadol) han debido ser comparados cara a cara en el mismo estudio, o, cuando menos, si se trata de estudios diferentes, la frecuencia basal de mal control del dolor con el tratamiento de referencia (en este caso, AINE) debería haber sido igual.

OTRAS MEDIDAS DE ASOCIACIÓN

Además de las medidas de asociación mencionadas hasta este punto, en ocasiones los investigadores emplean las dos siguientes:

Razón de prevalencias

A veces se conducen estudios de corte transversal para evaluar asociaciones. Dado que dichos estudios miden de manera simultánea la exposición al factor de interés y la presencia del desenlace, tan solo permiten obtener prevalencias de punto. Dicho de otra forma, como no hay seguimiento, esos estudios no pueden determinar la incidencia (el riesgo) del evento según la exposición, y, en consecuencia, no permiten obtener las medidas de asociación para comparar riesgos, descritas previamente (RAR, RR, etc.). En tales circunstancias, se estima la *razón de prevalencias*; es decir, la prevalencia hallada en el grupo expuesto dividida por la de los pacientes no expuestos. La interpretación del valor obtenido es similar a la descrita para el RR: un valor < 1,0 sugiere que la exposición se asocia a una reducción en la prevalencia del desenlace de interés, mientras que cifras iguales

a 1,0 o mayores implican que no hay asociación, o que la exposición coincide con un aumento de la prevalencia, respectivamente.

EJEMPLO

El primer estudio que propuso la asociación entre la presencia de *H. pylori* y la úlcera péptica fue un estudio de corte transversal conducido en Australia y publicado en 1984[16]. Los investigadores incluyeron a 100 pacientes remitidos para endoscopia digestiva alta, en quienes, durante el procedimiento, se estableció de forma simultánea la presencia de la bacteria y del desenlace. De los 31 pacientes con úlcera, 27 (87 %) tenían evidencia de la bacteria, que solo se encontró con endoscopia normal en 8 de los 16 sujetos (50 %). La razón de prevalencias fue de 1,74, lo cual indica que la presencia de *H. pylori* en los pacientes con úlcera fue 1,74 veces (o el 74 %) mayor que en los pacientes sin úlcera.

Razón de "peligros" (*hazard ratio* en inglés)

Algunas veces, los investigadores pueden estar interesados en considerar tanto la frecuencia con la que se presenta un desenlace como el tiempo hasta la ocurrencia del evento. En dichos estudios, llamados ***de supervivencia***, se mide el tiempo transcurrido para cada paciente desde el comienzo de la exposición hasta cuando se presenta el evento de interés (curación, recaída, muerte, etc.). Si el evento ocurre durante el tiempo de seguimiento, el tiempo de "supervivencia" se denomina ***exacto***; por el contrario, si el evento no se presenta durante el seguimiento, el tiempo de supervivencia es indeterminado, o *censurado*. Esos estudios permiten medir la tasa de "peligro" (***hazard rate*** en inglés) del evento; esto es, la probabilidad del evento en un intervalo dividido por la duración del intervalo; si tales intervalos son muy cortos, la tasa de "peligro" representa la tasa instantánea del evento en cualquier momento durante el tiempo de seguimiento[17]. Si se obtienen las tasas de "peligro" en un grupo de sujetos expuestos y en otro no expuesto, se puede obtener la razón de "peligros" (***hazard ratio*** en inglés); la interpretación del valor obtenido es similar a la descrita para las otras medidas relativas mencionadas en este capítulo (un valor de 1,0 sugiere que no hay asociación; cifras < 1,0 que la exposición reduce el riesgo del desenlace de interés, y cifras mayores, lo contrario).

EJEMPLO

En el curso de un experimento aleatorizado se asignó al azar a 170 342 sujetos entre 55 y 64 años de edad para ser sometidos o no a una sigmoidoscopia flexible para determinar si el procedimiento reducía la frecuencia de cáncer de colon y la mortalidad por dicha enfermedad[18]. Luego de una mediana de seguimiento de 11,2 años, se encontró que la exposición a la endoscopia redujo en el 23 % y en el 31 %, respectivamente, las tasas de incidencia y de mortalidad para cáncer de colon (razones de "peligro" de 0,77 y 0,69, respectivamente).

MEDIDAS DE PRECISIÓN

Aun si un estudio para calcular la frecuencia de un evento o su asociación a una exposición da resultados que parecen válidos (porque los métodos utilizados son apropiados) es necesario establecer la precisión de los hallazgos.

EJEMPLO

Un investigador quiere calcular la frecuencia de fumadores entre los 900 estudiantes de pregrado de una facultad de medicina. Para ello, decide elegir al azar, mediante un sorteo con todos los estudiantes, a un total de 100 alumnos, a quienes aplica una encuesta. En dicha muestra, el investigador observa una frecuencia de fumadores (en este caso, la prevalencia) del 0,25 % o el 25 %. Como la muestra fue seleccionada aleatoriamente, el investigador concluye que ésta debe de ser también la frecuencia de fumadores en los 900 alumnos.

Desafortunadamente, el investigador no está totalmente en lo cierto. Para ilustrar el punto, se puede pensar qué ocurriría si él repitiera muchas veces su estudio, y cada vez con una muestra aleatoria nueva de 100 de los 900 estudiantes. Esto equivale a seleccionar la muestra al azar, aplicar la encuesta, obtener la prevalencia, devolver a los 100 estudiantes del caso al grupo total y repetir múltiples veces el procedimiento. Lo más factible es que en cada oportunidad se obtenga una prevalencia de fumadores ligeramente distinta. Si este proce-

dimiento se repitiera *n* veces, al final el investigador tendría una serie de *n* prevalencias. Si cada prevalencia se considera una observación individual, se podría obtener la distribución de frecuencias para las prevalencias observadas. Esta distribución de las *n* observaciones presenta varias características interesantes. Quizá, la más importante es que el promedio de estas *n* prevalencias es la mejor aproximación a la verdadera prevalencia de la población; o sea, a la que se obtendría si se hubiera entrevistado a los 900 estudiantes[3].

No obstante lo anterior, en la vida real ningún investigador quiere obtener nuestras repetidas del mismo tamaño. Por fortuna, otra de las propiedades de esta distribución hipotética permite calcular a partir de una sola medición cuáles serían los valores obtenidos si el estudio se repitiera con muestras del mismo tamaño (los 100 estudiantes del ejemplo). Algunos procedimientos matemáticos simples permiten obtener los intervalos de confianza en torno a los valores concretos aportados por el estudio[19-21].

Se puede definir el intervalo de confianza (IC) del 95 % (IC 95 %) como el rango de valores dentro de los que el investigador tiene el 95 % de certeza de que está el verdadero valor del fenómeno que se pretende calcular en la población.

La fórmula para obtener el IC 95 % varía de acuerdo con la naturaleza de lo que se esté midiendo (prevalencia, incidencia, RR, etc.).

La **tabla 9.4** presenta las fórmulas necesarias para calcular el error estándar y los IC 95 %, para las medidas de frecuencia y de asociación descritas antes en este capítulo[19,22]. Como puede verse, la magnitud de los IC depende, fundamentalmente, de la variabilidad en la medición del fenómeno de interés, que corresponde al numerador de las fórmulas, y del tamaño de la muestra estudiada, que es el denominador. De esta forma, cuanto mayor sea la variabilidad o menor el número de individuos que componen la muestra, más amplio será el IC, y, por lo tanto, menor la precisión del cálculo.

[3] Esta propiedad es una consecuencia del llamado *teorema del límite central*.

EJEMPLO

Retomando el ejemplo de la frecuencia de fumadores, en el que se halló una prevalencia del 25 % al estudiar a 100 individuos, los límites de los IC 95 % se extienden del 16,5 % al 33,5 %. Ello significa que con este tamaño de muestra el investigador tiene el 95 % de certeza de que la verdadera frecuencia de fumadores en los 900 estudiantes de la facultad está en algún punto entre estos límites: el 16,5 % o el 33,5 %. Al aumentar la muestra de 100 a 400 alumnos, se obtiene un IC 95 % que va del 20,8 % al 29,5 %, bastante más preciso que el obtenido con solo 100 sujetos.

La **figura 9.3** muestra los riesgos relativos con sus IC 95 % correspondientes para varios de los desenlaces del experimento Magpie. La línea vertical central corresponde a un RR de 1,0, lo cual señala que no hay diferencias en la frecuencia del evento entre los grupos del estudio (gestantes tratadas con sulfato de magnesio o con placebo). Los valores ubicados a la izquierda corresponden a resultados donde la terapia experimental produce mejores efectos que el control, mientras los ubicados a la derecha indican lo contrario. El punto de cada desenlace corresponde al RR hallado en el estudio, y las líneas horizontales, al IC 95 % correspondiente.

El análisis de estas cifras y de la **figura 9.3** permite observar cada resultado y su precisión. El RR de eclampsia con el uso de sulfato de magnesio es de 0,42, con IC 95 % de 0,29-0,61. Según estos datos, se tiene el 95 % de certeza de que el verdadero RR para eclampsia asociado al uso de sulfato de magnesio está entre esos dos límites, que corresponden a la reducción en el riesgo de eclampsia del 71 % y del 39 %, respectivamente; como puede observarse, la línea que representa los IC 95 % del RR para este desenlace no se sobrepone con el valor de 1. Con base en dicho resultado, se puede concluir que la administración de sulfato de magnesio a mujeres con preeclampsia reduce el riesgo de desarrollar eclampsia.

Por otra parte, el RR para muerte materna es de 0,55, con IC 95 % de 0,26-1,14. Este intervalo, además de indicar menos precisión que para la eclampsia, no permite concluir que haya diferencia entre los tratamientos, pues los resultados son compatibles con una reducción del riesgo de morir al emplear sulfato de magnesio del 74 % (RR 0,26), así como con un aumento de este del 14 % (RR 1,14).

Tabla 9.4 Fórmulas para calcular los IC 95 % para medidas de frecuencia y de asociación, y sus resultados al aplicarlas al riesgo de eclampsia en el experimento Magpie

Medida	Fórmula	Experimento Magpie
Error estándar de una proporción	$ES_P = \sqrt{\frac{p(1-p)}{n}}$ p = proporción n = número de sujetos	$\sqrt{\frac{0{,}008(1-0{,}008)}{5.055}} = 0{,}0012$
IC 95 % de una proporción	$p \pm 1{,}96\ ES_P$	0,008 ± 1,96 × 0,0012 = 0,006 a 0,01
Error estándar de la RAR	$ES_{RAR} = \sqrt{\frac{pe(1-pe)}{ne} + \frac{pne(1-pne)}{nne}}$ *ne* y *nne* = número de expuestos y no expuestos	$\sqrt{\frac{0{,}008x0{,}992}{5.055} + \frac{0{,}019x0{,}981}{5.055}}$ = 0,002
IC 95 % RAR	$RAR \pm 1{,}96 ES_{RAR}$	**0,011 ± 1,96 × 0,002** = 0,007 a 0,015
Error estándar del RR	$ES\ln_{RR} = \sqrt{\frac{1}{re} + \frac{1}{rne} - \frac{1}{ne} - \frac{1}{nne}}$ *re* y *rne* = eventos observados en expuestos y no expuestos	$\sqrt{\frac{1}{40} + \frac{1}{96} - \frac{1}{5.055} - \frac{1}{5.055}}$ = 0,187
IC 95 % RR	$RR \pm 1{,}96 ES_{RR}$	**RR = 0,42; $\ln$ RR = –0,868** –0,868 ± 1,96 × 0,187 = –1,235 a –0,501 **e^x – 1,235 = 0,29 a e^x – 0,501 = 0,61**
IC 95 % RRR	**[1 – límite inf IC 95% RR]100 % y** [1 – límite sup **IC 95% RR]100 %**	[1 – 0,29] 100 % = 71 % y [1 – 0,61] 100 % = 49 %
Error estándar del RRI	$ES\ln_{RRI} = \sqrt{\frac{1}{re} + \frac{1}{rne} + \frac{1}{ne–re} + \frac{1}{nne–rne}}$	$\sqrt{\frac{1}{40} + \frac{1}{96} + \frac{1}{5.015} + \frac{1}{4.959}}$ = 0,187
IC 95 % RRI	$RRI \pm 1{,}96 ES_{OR}$	**RRI = 0,42; $\ln$RRI = –0,868** –0,868 ± 1,96 × 0,187 = –1,235 –0,501 **e^x – 1,235 = 0,29-e^x – 0,501 = 0,61**
IC 95 % NNT	**Recíproco de los IC 95 % RAR**	**1/0,015-1/0,007 = 67 a 143**

Fuente: Magpie[9].

Es pertinente discutir la correspondencia entre los **IC 95 %** y las ***pruebas de significación estadística*** **(o sea, las que miden los valores de *p* mediante los métodos convencionales de** pruebas de hipótesis). Esta ***p* corresponde a la** probabilidad de que la diferencia en los riesgos sea debida al azar, o, lo cual es lo mismo, de que, en realidad, no haya diferencia entre ellos. Si esta ***p* es pequeña, usualmente ≤ 0,05** (valor seleccionado arbitrariamente, pero de manera uniforme, por la comunidad científica internacional), se concluye cuán poco probable es que el azar explique las diferencias encontradas.

Desenlace	RR	IC 95%	p
Eclampsia	0,42	0,29 a 0,61	0,000003
Muerte materna	0,55	0,26 a 1,14	0,150
Muerte neonatal	1,02	0,92 a 1,14	0,739
Mortinato	0,96	0,84 a 1,10	0,577

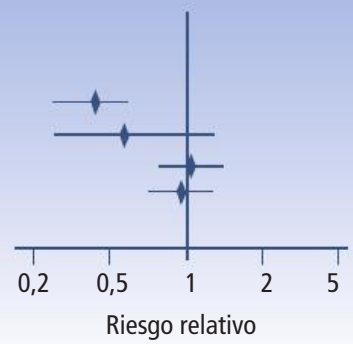


Figura 9.3 Riesgos relativos e IC 95% para algunos desenlaces del estudio Magpie.

EJEMPLO

Los riesgos de eclampsia en las mujeres tratadas con sulfato de magnesio y con placebo fueron del 0,8 % y el 1,9 %, respectivamente. La *p* para la comparación de los riesgos de muerte es de 0,0000028. Esta es la probabilidad de que la diferencia observada en los riesgos de muerte sea debida al azar. Como es poco probable que el azar explique los hallazgos, se puede inferir que, en realidad, hay una diferencia entre los riesgos asociada al uso del medicamento. Por su parte, la *p* para la asociación entre la intervención y la muerte materna **(figura 9.3)** es de 0,15, lo cual significa que el azar es una explicación probable para la diferencia observada.

Esta ***p*, no obstante ser muy popular entre** muchos investigadores, tiene grandes limitaciones[23-25]. En realidad, su valor no dice nada respecto a la magnitud o a la dirección de la diferencia encontrada, datos que son imprescindibles para tomar decisiones clínicas. El valor de la ***p* para eclampsia en el experimento Magpie** sería el mismo si los riesgos de eclampsia hubieran sido del 0,8 % con el placebo, y del 1,9 %, con el sulfato de magnesio; o sea, exactamente lo opuesto a lo hallado. La conclusión clínica, sin embargo, sería radicalmente distinta: la intervención aumentaría el riesgo de eclampsia, en lugar de reducirlo.

En contraste, y de gran utilidad para el clínico, además de informar sobre la significación estadística, los IC muestran la dirección y la precisión de la asociación de interés.

Finalmente, los IC 95 % informan cuándo un estudio que no ha hallado diferencias entre los grupos, conocido como ***estudio negativo*, posiblemente corresponda a un *verdadero negativo* o a un *falso negativo*** (para consultar las definiciones de estos términos, revisar el **capítulo 10** "Uso de pruebas diagnósticas en medicina clínica").

El experimento Magpie encontró un **RR para** mortalidad neonatal de 1,02, con **IC 95 % de** 0,92-1,14. El valor puntual es muy cercano a 1, y los límites del IC son bastante precisos, lo cual hace poco probable que el uso del sulfato de magnesio se asocie a grandes cambios en la mortalidad neonatal. Esto correspondería a un hallazgo verdadero negativo.

Una situación muy distinta sería hallar el mismo valor puntual de 1,02, pero con IC 95 % entre 0,42 y 1,54. En este caso, la baja precisión no permite descartar diferencias clínicamente importantes, pues estos resultados podrían ser compatibles con una reducción del 58 % en el riesgo del evento, así como con un aumento de este del 54 %. El estudio descrito es un estudio falso negativo, y sería más apropiado llamarlo "no conclusivo". La razón más común para tal tipo de resultados es bajo poder, secundario a tamaño de muestra insuficiente.

CALCULADORAS

Las fórmulas de las **tablas 9.2** y **9.4 permiten** calcular las medidas de asociación descritas a lo largo de este capítulo, así como sus IC. El procedimiento, sin embargo, puede resultar engorroso. Hay múltiples sitios en Internet a través de los cuales se puede acceder a calculadoras que simplifican el proceso. Más abajo se presentan los vínculos para algunos de dichos recursos. Casi

todos ellos presentan una tabla de 2 × 2 similar a la de la **tabla 9.1**, pero con las celdas vacías. Todo lo que el usuario debe hacer es ingresar en cada celda el valor observado y pinchar la tecla de comando para ejecutar los cálculos; en fracciones de segundo, la calculadora dará los resultados.

Hay, no obstante lo anterior, dos pasos cruciales en los que el computador no presta mayor ayuda. El primero es el ingreso de las cifras a la tabla. Como ya se dijo, es muy importante que el usuario tenga completa claridad respecto al valor que debe consignar en cada una de las celdas; el programa es incapaz de detectar errores en este paso, y hará los cálculos independientemente de si los valores consignados son correctos o erróneos. El segundo paso es la interpretación de los resultados, que es completa responsabilidad del usuario. Aquí, de nuevo, el computador no presta ninguna ayuda. Los lectores que quieran familiarizarse con el proceso pueden utilizar las cifras del experimento Magpie que se presentan en la **tabla 9.1** para practicar los procedimientos empleando estas calculadoras:

- 2-way Contingency Table Analysis (http://statpages.org/ctab2x2.html)
- OpenEpi. Open Source Epidemiologic Statistics for Public Health" (http://www.openepi.com/Menu/OE_Menu.htm)
- Evidence Based Calculator (http://moosenose.com/EBCalculator.htm)
- "EpiMax Table Calculator (http://www.healthstrategy.com/epiperl/epiperl.htm)

Esta medida corresponde a una familia de mediciones conocidas como fracciones atribuibles, y, específicamente, la fracción de exceso de casos[4], y que, a su vez, en el caso del **RR** se puede denominar cambio del efecto relativo, el cual puede ser positivo o negativo y se expresa como porcentaje. Cuando es positivo, se denomina exceso relativo de riesgo, y cuando es negativo es **RRR**. En general, se computa sustrayendo 1 a la razón de medidas de frecuencia relativa y multiplicando esta diferencia por 100. Por ejemplo, en el caso de un estudio de cohorte de factores de riesgo que muestre que la incidencia acumulada de infarto de miocardio es 4 veces mayor en pacientes fumadores (expuestos) que en no fumadores (no expuestos), la razón entre las dos incidencias, o **RR**, es 4:1, o, simplemente, 4. ¿En qué magnitud se incrementó el riesgo en los fumadores? En (**RR**-1) × 100, o (4-1) × 100 = 300 %. Multiplicar por 4 el riesgo (la frecuencia esperada) es lo mismo que incrementarlo en el 300 % (el 100 % basal, más el 300 % adicional da al final el 400 %, o 4 veces más). En general, para calcular la **RRR** se puede usar la misma fórmula, donde se sustrae 1 del **RR** y se multiplica por 100. Se debe tener en cuenta que el signo negativo de la sustracción indica que se redujo el "exceso de riesgo", o que el cambio relativo del riesgo fue negativo. Por tal motivo, en la literatura metodológica sobre el cálculo de la **RRR** (particularmente, en experimentos que evalúan la efectividad de intervenciones, y en los cuales se espera que la exposición disminuya el riesgo; es decir, que el **RR** sea < 1 se acostumbra invertir la fórmula: (1−**RR**) × 100; y si la diferencia es positiva, se interpreta como la magnitud de la RRR, tal como se muestra en la **tabla 9.2**, en vez del confuso "exceso negativo" de riesgo.

Referencias

1. Porta M. A dictionary of epidemiology. 5th ed. Nueva York: Oxford University Press; 2014.
2. Fletcher RH, Fletcher SW, Fletcher GS. Frequency. En: Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology. The Essentials. 5th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2014. p. 17-30.
3. Gordis L. Measuring the occurrence of disease. En: Gordis L. Epidemiology. 4th ed. Philadelphia: Saunders Elsevier; 2009. p. 37-57.
4. Greenland S, Rothman KJ. Measures of occurrence. En: Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2008. p. 32-50.
5. Morillo LE, Díaz J, Estévez E, et al. Prevalence of erectile dysfunction in Colombia, Ecuador and Venezuela: a population-based study (DENSA). Int J Impot Res. 2002;14:S10-8.
6. Efird M, Rojas MA, Lozano JM, et al. Epidemiology of nosocomial infections in selected intensive care units in Colombia, South America. J Perinatol. 2005;25:531-6.
7. Fletcher RH, Fletcher SW, Fletcher GS. Risk: basic principles. En: Fletcher RH, Fletcher SW, Fletcher

GS. Clinical epidemiology. The Essentials. 5th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2014. p. 50-60.
8. Richardson WS, Wilson MC. The process of diagnosis. En: Guyatt G, Rennie D, Meade MO, et al, editores. Users' Guides to the medical literature: A manual for Evidence-Based Clinical Practice. 2nd ed. New York: McGraw-Hill; 2008. p. 169-78.
9. Altman D, Carroli G, Duley L, et al. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: a randomised placebo-controlled trial. Lancet. 2002;359:1877-90.
10. Jaeschke R, Guyatt G, Barratt A, et al. Does treatment lower risk? Understanding the results. En: Guyatt G, Rennie D, Meade MO, et al, editores. Users' Guides to the medical literature: A manual for Evidence-Based Clinical Practice. 2nd ed. New York: McGraw-Hill; 2008. p. 113-26.
11. Lapaucis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318:1728-33.
12. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analysis of randomized trials. J Clin Epidemiol. 1994;47:881-9.
13. Straus SE, Glasziou P, Richardson WS, et al. Therapy. En: Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine. How to practice and teach it. 4th ed. Edinburgh: Churchill Livingston Elsevier; 2011. p. 67-136.
14. Sedgwick P. Odds and odds ratios. BMJ. 2013;347:f5067.
15. McAlister FA. The "number needed to treat" turns 20—and continues to be used and misused. Can Med Assoc J. 2008;179:549-53.
16. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet. 1984;1:1311-5.
17. Sedgwick P. Hazard and hazard ratios. BMJ. 2012;345:e5980.
18. Atkin WS, Edwards R, Kralj-Hans I, et al. UK flexible sigmoidoscopy trial investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet. 2010;375:1624-33.
19. Gardner JM, Altman DG. Intervalos de confianza y no pruebas p: estimación en vez de pruebas de hipótesis. Bol Of Sanit Panam. 1993;114:536-49.
20. Sim J, Reid N. Statistical inference by confidence intervals: issues of interpretation and utilization. Phys Ther. 1999;79:186-95.
21. Sedgwick P. Confidence intervals: predicting uncertainty. BMJ. 2012;344:e3147.
22. Altman DG. Appendix 1: Confidence Intervals. En: Sackett DL, Richardson WS, Rosemberg W, et al. Evidence-Based Medicine: How to practice and teach EBM. 3rd ed. Edinburgh: Churchill Livingston Elsevier; 2005. p. 263-77.
23. Kain ZN, MacLaren J. P less than .05: What does it really mean? Pediatrics. 2007;119:608-10.
24. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. Austr Crit Care. 2010;23:93-7.
25. Sedgwick P. Confidence intervals and statistical significance: rules of thumb. BMJ. 2012;345:e4960.

CAPÍTULO

10

Uso de pruebas diagnósticas en medicina clínica

Álvaro J. Ruiz Morales

INTRODUCCIÓN

En el proceso de práctica de la medicina clínica, permanentemente se deben tomar decisiones; casi siempre, con bases que conllevan una alta dosis de incertidumbre. Riegelman[1] describió el diagnóstico médico como un "intento de tomar decisiones adecuadas a pesar del uso de información inadecuada". En un intento por reducir la incertidumbre inherente al manejo de seres vivos, se usan pruebas diagnósticas, cuya función es no solo reducir la incertidumbre, sino cuantificar y describir las dudas que puedan quedar.

La rápida evolución de la tecnología y la aparición de nuevas técnicas diagnósticas han hecho perder un poco el concepto de que la fase inicial del proceso diagnóstico empieza siempre con el interrogatorio y con al examen físico. Ambos no solo han sido relegados a un segundo plano, sino que, en muchos casos, no se los considera como pruebas diagnósticas. Y, sin embargo, se ha mostrado que hasta en el 56 % de los casos el diagnóstico puede ser hecho solamente con el interrogatorio, y hasta en el 73 %, con el interrogatorio y el examen físico[2].

El *interrogatorio* y la *historia clínica* son también herramientas diagnósticas, y deben ser consideradas como tales; debe reconocerse que tienen sensibilidad y especificidad, y que dependen, en gran medida, del operador[3].

La evolución de la ciencia ha traído consigo la disponibilidad, cada vez mayor, de pruebas diagnósticas, con frecuencia complejas y costosas. La falta de formación y de información sobre las pruebas, sobre sus características y sobre la eficiencia de su uso ha llevado a una difícil situación de los sistemas de salud, que necesita una racionalización urgente.

Se debe decidir en cada caso si se justifica o no hacer exámenes diagnósticos, cuál o cuáles practicar, cuándo hacerlos, si se los hace además de otras pruebas; o, en su lugar, cuál es el orden adecuado de los exámenes que se practicarán, y cómo se interpretarán los resultados[4]. Como ejemplos, no se justifica la auscultación rutinaria de la región lumbar en busca de estenosis de la arteria renal[5], y ante la sospecha de encefalitis por herpes la resonancia magnética ofrece mejores características operativas que la tomografía[6]. El clínico debe incorporar el conocimiento profundo sobre las pruebas que puede usar en su desempeño.

LA PRUEBA DIAGNÓSTICA IDEAL

En la interpretación de una prueba diagnóstica hay algunas precauciones que siempre deben tenerse en cuenta. Como se verá en el ejemplo, no siempre son evidentes las bondades o los errores de un método diagnóstico.

EJEMPLO

En un hospital universitario se tomó, por razones logísticas, la decisión de fotocopiar la plantilla transparente, proporcionada por el fabricante, que se debe usar para la determinación, utilizando la radiografía, del tamaño apropiado de la prótesis de cadera que necesita cada paciente. Con sorpresa, se observó que en todas las cirugías se encontraba que el tamaño de las prótesis era inapropiado: siempre menor que el necesario. No valieron las recomendaciones de cuidado ni el reentrenamiento en el uso de la plantilla, pues se repetía la situación anómala. Hasta cuando un ortopedista de mente crítica encontró que, en el proceso, la señora encargada de las fotocopias decidió (pues el original superaba el mayor tamaño disponible de acetatos) tomar las copias *haciendo una reducción del 15%*.

Hay dos condiciones que debe cumplir una prueba. La primera de ellas, la ***repetibilidad***; esto es, que pueda ofrecer resultados comparables en situaciones similares. Pero, como se ilustró en el ejemplo anterior, el hecho de que se repita el valor de una medición no la hace verdadera. Aún si se hubieran repetido las evaluaciones muchas veces, el resultado habría sido el mismo. Y el clínico debe tener claro que el que algo se repita no significa que sea cierto.

Es necesaria otra condición: el valor entregado por la prueba debe ajustarse a la realidad. A esta característica de una prueba se la llama *validez*, y es claro que no tiene relación con la repetibilidad. También lo es que las dos deben coexistir en una prueba para que se la pueda aceptar. La estrategia diagnóstica debe ser válida (es decir, proporcionar el valor correcto), pero debe tener también repetibilidad (la capacidad para ofrecer resultados comparables en condiciones similares).

La **figura 10.1** muestra el concepto de la necesidad de validez aun si hay repetibilidad. En el caso del arquero ***a***, hay repetibilidad, pero también hay un sesgo, ya que hay una ***desviación sistemática***; es decir, un error que ocurre siempre

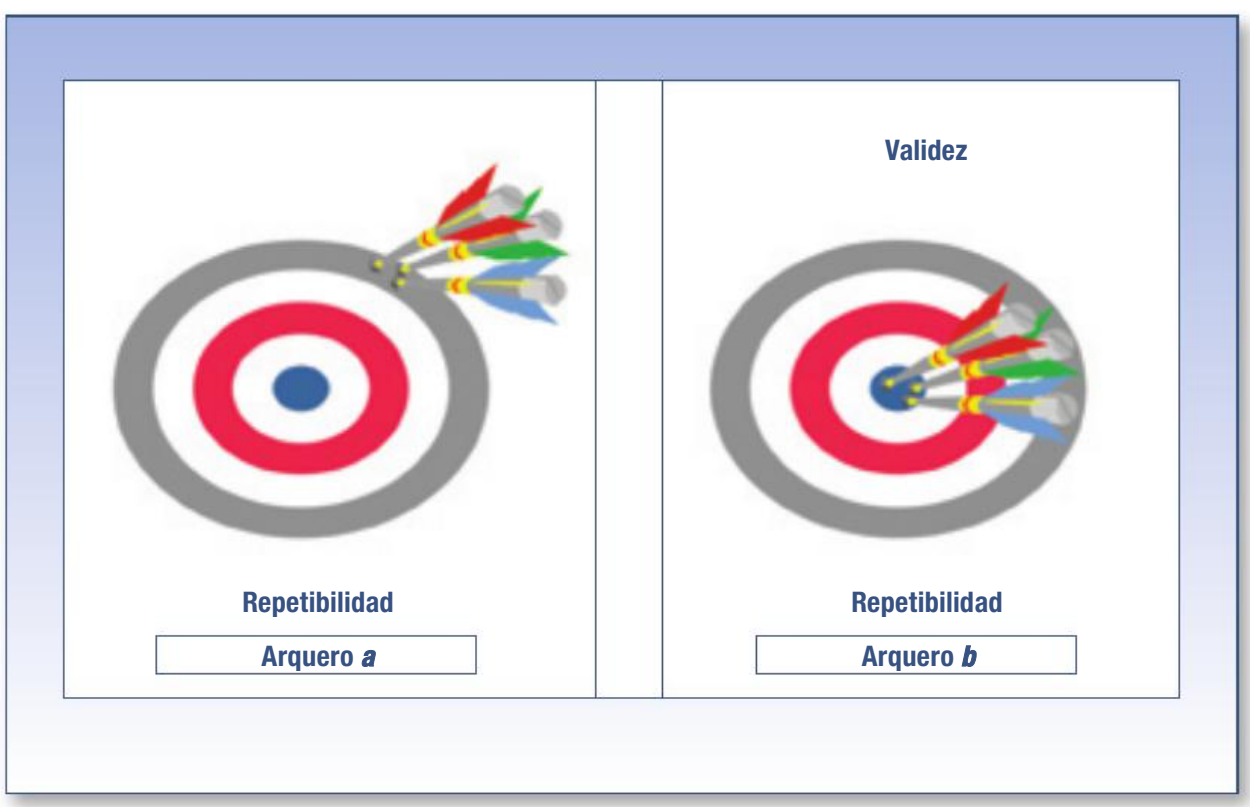


Figura 10.1 Condiciones para una prueba: repetibilidad y validez.
Fuente: Elaboración propia.

en el mismo sentido. El arquero *b* tiene también repetibilidad (algunas veces, llamada ***precisión***), pero, además, tiene validez.

El principio fundamental sobre el que descansan las pruebas diagnósticas es que los individuos enfermos son diferentes de los individuos sin la enfermedad, y que dichas pruebas pueden hacer esa distinción. Una prueba diagnóstica perfecta necesitaría, según Riegelman[1], tres condiciones fundamentales: 1) que todos los enfermos tengan el mismo resultado, 2) que los no enfermos tengan, a su vez, el mismo resultado, y 3) que los posibles resultados sean solamente *enfermo* o *no enfermo*. El **cuadro 10.1** recuerda las características de la prueba diagnóstica perfecta.

Usos de las pruebas diagnósticas

Por supuesto, lo enunciado sobre prueba diagnóstica perfecta no se cumple en la vida real. Con frecuencia, la información que se tiene en un caso particular no es suficiente para formular un diagnóstico ni para tomar decisiones. En este caso, debe recurrirse a una prueba diagnóstica, y es raro que las pruebas cumplan a cabalidad las condiciones ideales. La decisión de usar las pruebas, incluso a sabiendas de que no son perfectas, depende del uso que se le vaya a dar a la información y del objetivo en la búsqueda. Se reconocen cinco usos básicos de las pruebas diagnósticas, los cuales se describen en el **cuadro 10.2**.

El nivel más básico de exámenes es el *tamizado*, o búsqueda de una situación en una población de baja probabilidad. La toma generalizada de presión arterial en los centros comerciales, los aeropuertos y los sitios públicos es un buen ejemplo. En este caso, es importante encontrar casos (o, al menos, a sospechosos) y la prueba debe tener la menor cantidad posible de resultados falsos negativos. En el tamizado para hipertensión en una estación de buses, es importante encontrar a personas con valores elevados, y si hay algunos valores falsos positivos, las consecuencias no son severas, ya que siempre se remite a estas personas para una segunda evaluación. Un resultado falso negativo, sin embargo, daría confianza al individuo, y puede demorar mucho tiempo (incluso, varios años) la posibilidad de detectar la enfermedad.

Una prueba de tamizado es útil cuando la condición tiene una prevalencia relativamente alta en la fase preclínica, cuando la enfermedad es seria y cuando el resultado es mejor si se inicia el tratamiento antes de que sea aparente la condición.

Los programas de tamizado para la hipertensión arterial[7] y para el cáncer de cuello uterino[8] son buenos ejemplos de condiciones en las que se justifica la práctica rutinaria. En el caso de los cálculos en la vesícula, sin embargo, no hay indicación para intervención cuando son asintomáticos, y la cirugía, cuando es necesaria, no conlleva complicaciones graves, por lo que no se justifica su búsqueda en personas sin síntomas.

Si la condición buscada es muy rara, el proceso es muy ineficiente y no está justificado (por ejemplo, la ya mencionada auscultación rutinaria en busca de soplos por estenosis renal). Hay controversia en cuanto a la utilidad de algunas pruebas de tamizado, como el antígeno prostático específico (**PSA**, por las siglas en inglés de ***prostate specific antigen***)[9], para la búsqueda de cáncer prostático. El clínico y el tomador de decisiones deben conocer la utilidad de cada programa de tamizado, para tomar la decisión correcta en cuanto a su recomendación rutinaria.

Un segundo nivel de exámenes, más avanzado, se presenta cuando se busca una condición en sujetos con factores de riesgo; es decir, con aumento de la probabilidad de sufrir la enfermedad. Un examen de glucemia en un sujeto obeso y con antecedentes de diabetes en la familia corresponde a esta indicación, la *búsqueda de casos*. En este caso, también es

CUADRO 10.1 Características de la prueba diagnóstica ideal

1. Las personas con la enfermedad tienen, todas, resultado uniforme en el examen.
2. Las personas sin enfermedad tienen, todas, resultado uniforme y diferente del anterior.
3. Todos los resultados del examen pertenecen o bien a la categoría de personas con enfermedad, o bien, a la categoría de aquellos sin enfermedad.

CUADRO 10.2 Indicaciones para el uso de las pruebas diagnósticas

- Tamizado
- Búsqueda de casos
- Descarte de un diagnóstico
- Confirmación de un diagnóstico
- Seguimiento

preferible tener una prueba que ofrezca un bajo número de resultados falsos negativos, y el programa de tamizado ***se justifica*** por la posibilidad de encontrar una condición seria, modificable y que acarrea serias consecuencias si no se la diagnostica tempranamente, y, a conciencia, se acepta que pueda haber algunas personas que al ser sometidas a una prueba de confirmación, por haber sido positivas en el tamizado, no tengan la enfermedad.

En algunos casos, se quiere estar seguro de que una condición específica no está presente. En tales casos, la modalidad es el ***descarte de un diagnóstico***. Se requiere que haya baja probabilidad preexamen, pero que las consecuencias, en caso de que exista la enfermedad, sean potencialmente serias.

EJEMPLO

Un paciente de 48 años de edad con antecedentes de reflujo gastroesofágico muy sintomático consulta por dolor retroesternal súbito intenso, irradiado a la región cervical. No hay hallazgos significativos en el examen físico. La principal probabilidad diagnóstica se relaciona con enfermedad acidopéptica por reflujo, y con que el dolor se deba a un espasmo esofágico. Sin embargo, se debe tener certeza de que no se trata de angina o de un infarto agudo de miocardio (IAM). La toma de un electrocardiograma y la medición de marcadores como troponina corresponden al uso de pruebas para descartar un diagnóstico.

En el mismo ejemplo, una endoscopia digestiva que muestre esofagitis aguda puede demostrar con certeza el diagnóstico. Este uso, llamado ***confirmación de un diagnóstico***, debe emplearse de manera juiciosa. Puede ser muy útil y resolver dudas de manera definitiva, así como facilitar el enfoque terapéutico y establecer el pronóstico. Pero también puede ser innecesario, y el gasto en la atención en salud puede dispararse de manera dramática en manos inexpertas. Un viejo aforismo médico dice que "cuando no es necesario hacer un examen, es necesario no hacer el examen". Una manera importante de ayudar a la decisión de toma de una prueba es la pregunta: ***¿qué se hará con el resultado del examen?*** Si la respuesta a dicha pregunta es igual para un resultado positivo y para un resultado negativo, ***no hay ninguna utilidad en haber hecho el examen***.

Para la decisión de hacer o no el examen, es útil el concepto ***umbral de incertidumbre***, que se ilustra en la **figura 10.2**.

En el extremo izquierdo de la **figura 10.2** se representa la situación en la que la probabilidad del diagnóstico es ninguna y no hay duda al respecto, como en el caso de embarazo en una persona sin actividad sexual. En el otro extremo, la certeza absoluta existe y no cabe la duda, por lo cual tampoco se justifica llevar a cabo pruebas, si el objeto es determinar si está presente o no la condición sobre la que no hay incertidumbre.

En el intermedio se encuentra el rango amplio de niveles de certidumbre o incertidumbre, en el cual se justifica hacer pruebas adicionales. En algunos casos, aun en presencia de valores bajos de probabilidad, no se justifica hacer pruebas diagnósticas. En otros casos, y según el diagnóstico, los riesgos de no hacer el diagnóstico y los costos del examen llevarán a que se necesiten niveles más altos de probabilidad antes de tomar la decisión de hacer el examen.

EJEMPLO

Una mujer sana, de 17 años y sin factores de riesgo cardiovascular se presenta al servicio de urgencias con un dolor punzante en el hemitórax izquierdo, de aparición subaguda. No hay hallazgos significativos en el examen físico, ni antecedentes de importancia. La probabilidad de un IAM es tan baja que puede considerarse nula, por lo cual no se justifica la realización de exámenes especializados ni siquiera para descartar el diagnóstico.

Finalmente, las pruebas diagnósticas se utilizan para el ***seguimiento de casos***. Bien sea buscando asegurarse de que ha desaparecido la enfermedad (repetición de un urocultivo luego del tratamiento de una infección urinaria), o bien, para ver la evolución con el tratamiento (glucemia en el diabético en tratamiento), para controlar la aparición o el desarrollo de una reacción adversa (creatinina sérica en quien recibe aminoglucósidos), para seguir la actividad de la enfermedad (niveles de complemento en una enfermedad autoinmune), etc.

Características operativas de una prueba diagnóstica

En todas las indicaciones debe tomarse la decisión consciente de si se justifica el examen o no,

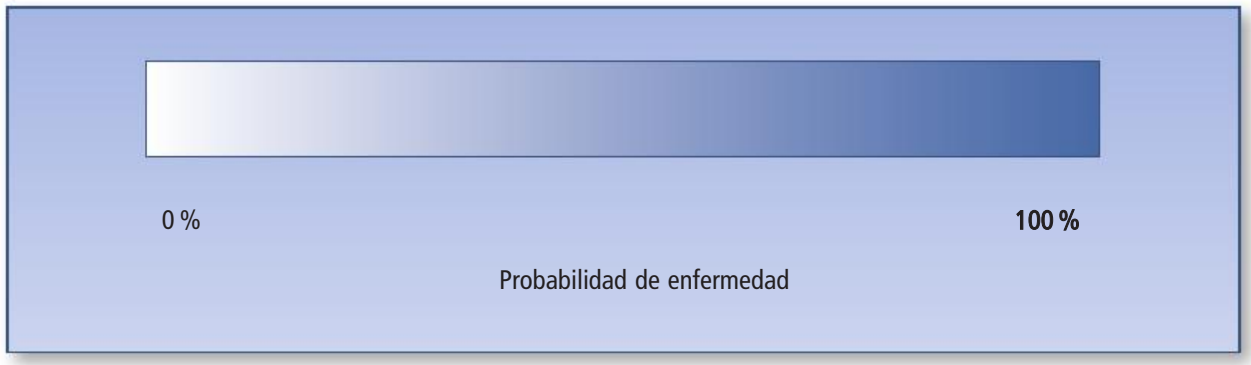


Figura 10.2 Umbral de incertidumbre.

cuándo hacerlo y, sobre todo, cuáles deben ser las características del examen, según los conceptos que se verán más adelante en este capítulo. Ya se mencionaron dos atributos necesarios para una prueba diagnóstica: la repetibilidad y la validez. En la evaluación de esta última se han descrito dos medidas de validez, a las que se ha llamado *características operativas*, y que tienen dos indicaciones precisas:

1. Se quiere evaluar una nueva prueba diagnóstica.
2. Se quiere interpretar el resultado de una prueba en un paciente concreto.

En el primer caso, se ***califica*** la prueba. Esta actividad la desarrolla quien hace investigación sobre pruebas diagnósticas, y no es habitual que el clínico corriente tenga dicha responsabilidad. Su objetivo es evaluar el desempeño de la nueva prueba y decidir si se justifica incluirla en el arsenal diagnóstico. En el segundo caso, están todos los médicos clínicos cada vez que reciben un resultado de una prueba diagnóstica.

Revisaremos inicialmente la primera situación, durante la cual se evalúa la prueba. Debe recordarse que esta actividad califica al examen, y no es una responsabilidad corriente del clínico, si bien este debe conocer muy bien los conceptos involucrados.

EVALUACIÓN DE UNA NUEVA PRUEBA DIAGNÓSTICA

El concepto fundamental de una prueba diagnóstica estriba en que con ella se busca diferenciar a quien tiene una condición de quien no la tiene. Y para hacer la evaluación de una prueba diagnóstica debe saberse si su resultado es indicativo de enfermedad en una persona en quien sabemos con certeza que existe la condición. Es indispensable tener una población en la que haya seguridad absoluta (o tanta como sea posible) de la existencia de la enfermedad, para aplicar la prueba y saber cuál es su desempeño. Si la prueba es positiva en quien tiene la enfermedad, esto corresponde a un ***verdadero positivo***, en tanto que si es negativa, ello corresponde a un *falso negativo*.

Al aplicar la prueba a la población en la que se tiene certeza, por el patrón de oro, de que no existe la condición, se espera que la prueba evaluada dé resultado negativo, a lo que se llamará resultado *verdadero negativo*; pero si es positiva, corresponde a un ***falso positivo***. Como puede verse, solamente al comparar con un patrón de oro puede hacerse la evaluación de la nueva prueba.

Al revisar los posibles resultados de una prueba, se ve cómo en un buen porcentaje de los casos los resultados pueden ofrecerse de forma dicótoma: positivo o negativo, sí o no, presente o ausente. En otros casos, hay opciones diferentes, como resultados positivos, dudosos o negativos; o calificaciones con probabilidad alta, intermedia o baja, o resultados por categorías numéricas, etc. Se revisará ahora la evaluación de pruebas con resultados dicótomos o aquellas en las cuales se establece un punto de corte.

Selección del patrón de oro

La primera condición para evaluar una nueva prueba es que exista un método confiable para hacer el diagnóstico que intenta la prueba. No se trata de un razonamiento circular, pero, ciertamente, es una situación difícil en medicina. Para poder evaluar una nueva prueba que diagnostique la condición X, debe tenerse un método diagnóstico muy preciso y confiable para esa

misma condición. Al examen que da la certeza de la condición se le llama ***estándar de oro***, o *patrón de oro*. Habitualmente, el patrón de oro es un examen de tejido (biopsia o autopsia) o una prueba difícil o costosa que no se utiliza de manera rutinaria, o una compleja combinación de criterios.

Es vital que la selección del patrón de oro sea adecuada, puesto que la evaluación de la prueba nueva será completamente inadecuada si no se tiene certeza de la presencia o la ausencia de la enfermedad. La medición de filtración glomerular no es práctica en el ejercicio corriente de la medicina, por lo cual se la reemplaza por la medición de la depuración de creatinina, que, si bien tiene características operativas ligeramente menos aceptables, es más fácil de hacer, más económica y más práctica.

Medición de sensibilidad

Uno de los aspectos que se quiere evaluar es la capacidad para clasificar correctamente como enfermo al enfermo. Para evaluar esta característica operativa, se utiliza el patrón de oro con el fin de seleccionar una población de individuos que tenga, con certeza, la condición. La **figura 10.3** muestra el procedimiento que debe seguirse.

Una vez seleccionado un número de sujetos con certeza de la condición, según el patrón de oro (100, en el ejemplo), se aplica la prueba nueva y se clasifica a los sujetos, según el nuevo resultado, como ***positivos*** o ***negativos***. Se obtiene, entonces, un grupo de sujetos que sí están enfermos (por el patrón de oro) y que son identificados por la nueva prueba como positivos. A este grupo se le llama ***verdaderos positivos***; el cálculo de la *tasa de verdaderos positivos* muestra la capacidad de la prueba para clasificar correctamente al enfermo como enfermo. A esta característica se la llama *sensibilidad*.

Si en la **figura 10.3** el número de pacientes correctamente clasificados fuera de 92, la sensibilidad de la prueba sería del 92 %. Se interpreta este resultado así: la nueva prueba clasifica correctamente como enfermos a 95 de cada 100 enfermos. Si el número de sujetos elegidos inicialmente no es 100, se obtiene el porcentaje de clasificación correcta mediante una regla de tres. Habitualmente, se asignan letras a las casillas de la tabla, que para la columna de enfermos son la ***a*** en la casilla de verdaderos positivos, y la ***c***, en la de falsos negativos. De manera que el cálculo de la sensibilidad, si se quisiera representar en una fórmula, sería:

$$Sensibilidad = a/(a+c)$$

La sensibilidad puede definirse como la capacidad de la prueba para clasificar correctamente al enfermo como enfermo, o como la probabilidad de tener un resultado positivo si se tiene la enfermedad.

A la casilla de resultados negativos en este grupo de la **figura 10.3** se la llama falsos negativos; también puede calcularse una tasa de *falsos negativos*, que en el ejemplo anterior sería del 8 %.

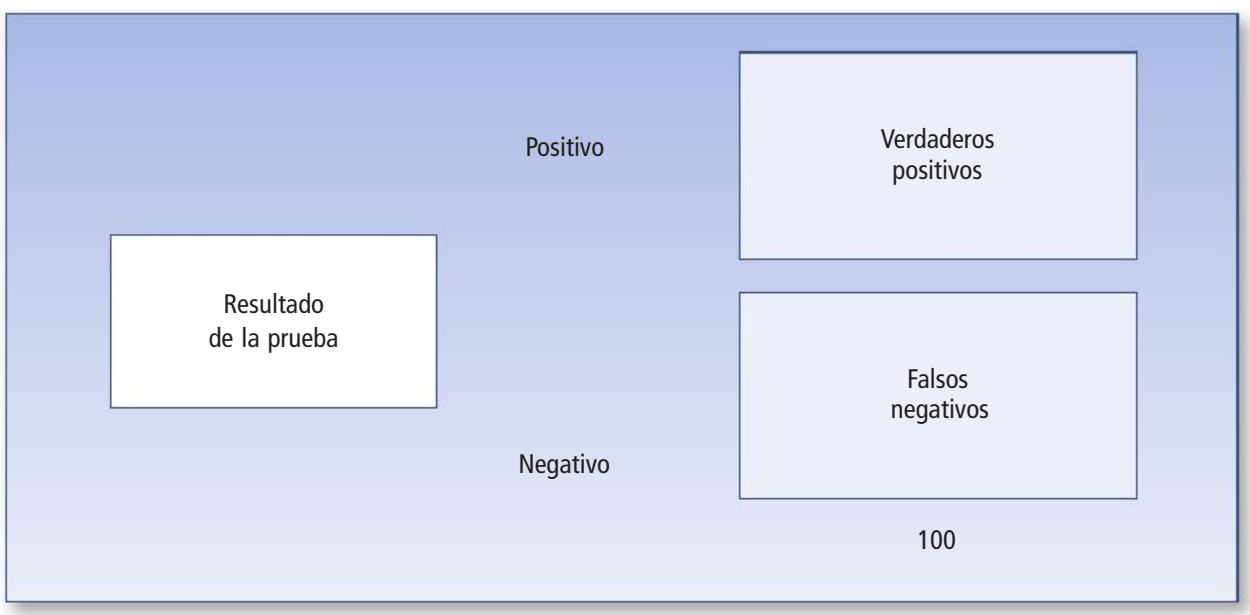


Figura 10.3 Evaluación de sensibilidad.

Puede verse cómo una prueba muy sensible tiene pocos falsos negativos, y viceversa. La utilización de una prueba muy sensible nos sugiere que un resultado negativo, probablemente, sea un resultado verdadero, pues tiene una enorme capacidad para detectar a los enfermos como enfermos y, por lo tanto, una baja tasa de falsos negativos.

MEDICIÓN DE ESPECIFICIDAD

Así como existe la duda de si el examen puede clasificar correctamente al enfermo, también existe la pregunta de si puede identificar al sujeto sano como sano. Para determinar esta importante característica operativa, se selecciona un grupo de sujetos que estén libres de la enfermedad, según el patrón de oro, y se les aplica la prueba nueva. La **figura 10.4** muestra cómo se hace dicha evaluación.

En esta evaluación, el aspecto más importante es que la prueba pueda clasificar correctamente al sujeto sano como sano, lo que corresponde a los ***verdaderos negativos***. Si la prueba clasifica correctamente a 98 de los sujetos sanos, se diría que la capacidad de la prueba para identificar al negativo como negativo, o ***especificidad,*** es del 98 %. Otra interpretación es que la probabilidad de ser negativo en la prueba, si se es sano, es del 98 %. La ***tasa de verdaderos negativos,*** o ***especificidad***, es, entonces, el porcentaje del total de sujetos sanos que tiene resultado negativo. Un examen con alta especificidad es muy útil cuando el resultado es positivo, pues la tasa de falsos positivos (***complemento de la especificidad***) en estos casos es muy baja.

SENSIBILIDAD Y ESPECIFICIDAD

La presentación habitual para determinar las pruebas se hace a través de la ***tabla tetracórica***, o de cuatro casillas, con la asignación de una letra a cada casilla. La **figura 10.5** muestra la tabla que sirve de base a las fórmulas empleadas en la determinación de la sensibilidad y la especificidad, la cual reúne las dos columnas, previamente presentadas, de resultados en población con la enfermedad y resultados en población sin la enfermedad. La presentación a través de la tabla facilita los cálculos y la visión global, pero la presentación aislada del cálculo de resultados en los sujetos enfermos y de los resultados en los sanos facilita comprender los conceptos.

Para el cálculo de las características mencionadas, se utilizan las letras de la figura, como ya se mencionó, las cuales se han definido internacionalmente y se usan siempre de la misma

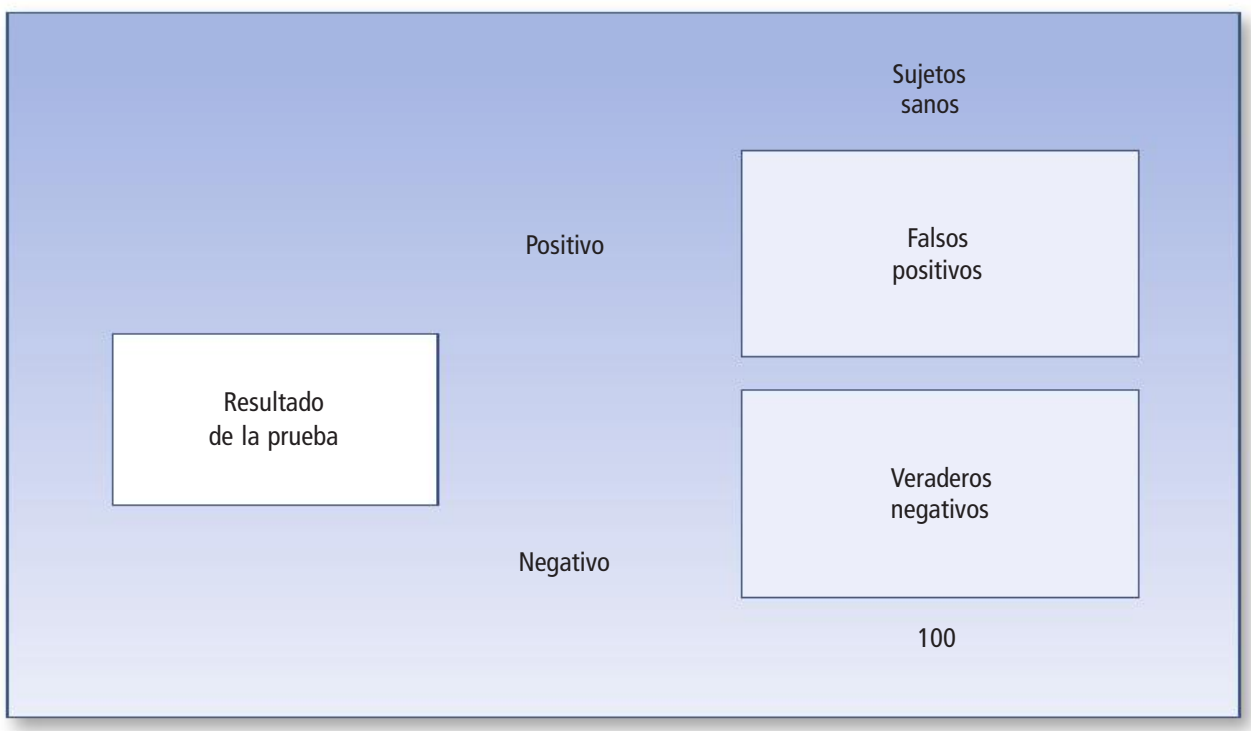


Figura 10.4 Evaluación de especificidad.

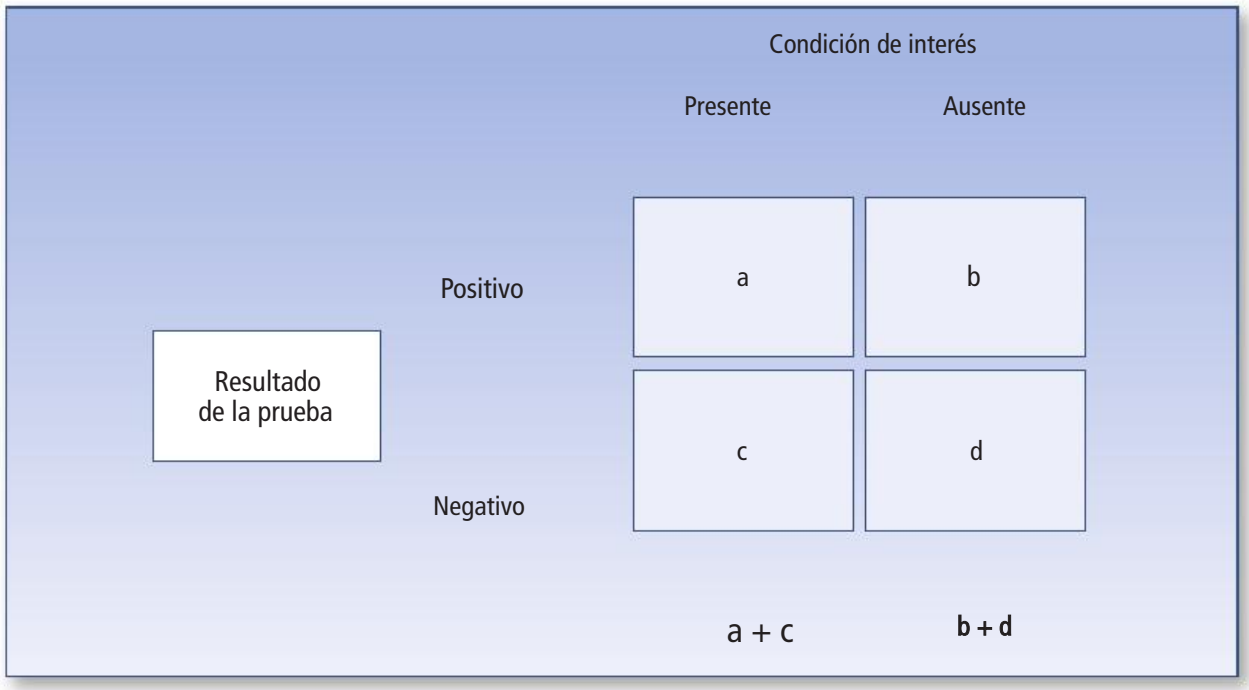


Figura 10.5 Resultados de la nueva prueba según el patrón de oro.

manera. Para los cálculos se usan las fórmulas siguientes:

1) para conocer el comportamiento de la prueba en personas con la enfermedad:

$$Sensibilidad = \frac{a}{a+c};$$

2) para conocer el comportamiento en personas sin la enfermedad:

$$Especificidad = \frac{b}{b+d};$$

más aún, hay una manera de reunir (con utilidad teórica) los dos aspectos en una tercera característica: la ***exactitud***, que corresponde al porcentaje de sujetos que resultaron correctamente clasificados por la prueba (tanto verdaderos positivos como verdaderos negativos):

$$Exactitud = \frac{a+d}{a+b+c+d}$$

INDICACIONES DE LAS PRUEBAS SEGÚN SENSIBILIDAD Y ESPECIFICIDAD

Según los conceptos anteriores, cuando sea muy riesgoso no detectar una enfermedad o cuando haya un precio alto que pagar si no se hace el diagnóstico, se debe escoger una prueba de alta sensibilidad. Estas pruebas son más útiles cuando sus resultados son negativos, pues la alta sensibilidad significa una baja tasa de falsos negativos. En las etapas iniciales de un estudio a un paciente, en los casos de tamizado (y, principalmente, cuando la probabilidad de la condición sea baja y el propósito sea descubrirla), se debe utilizar una prueba con ***alta sensibilidad***. Este es el caso de los exámenes periódicos de salud.

Por el contrario, cuando haya altos costos o riesgos (incluidos los riesgos emocionales) en un resultado falso positivo, deben emplearse pruebas con alta especificidad. Las pruebas muy específicas tienen mayor utilidad cuando sus resultados son positivos, pues tienen una baja tasa de falsos positivos. Se usan, principalmente, para la confirmación de diagnósticos.

RELACIÓN ENTRE SENSIBILIDAD Y ESPECIFICIDAD

Es ideal tener una prueba altamente sensible y altamente específica; y aunque, de hecho, existen, no son las más comunes. La prueba ELISA, para el virus de inmunodeficiencia humana (VIH), por ejemplo, tiene una sensibilidad y una especificidad cercanas al 100 %[10]. En un análisis metaanalítico de pruebas para virus de

la hepatitis B (HBV), se encontraron una sensibilidad del 98,07 % y una especificidad del 99,56 %[11]. En otras pruebas, sin embargo, los resultados no son tan buenos. En un metaanálisis de 132 estudios se reportó que la sensibilidad del electrocardiograma para la detección de isquemia durante el ejercicio fue del 68 %, y la especificidad, del 77 %[12]. Un estudio en 1997 reportó que para el diagnóstico de insuficiencia cardíaca congestiva (ICC), la presencia de S_3, o ***ritmo de galope***, tenía una sensibilidad del 24 % y una especificidad del 99 %, en tanto que la disnea de ejercicio tenía una sensibilidad del 100 %, pero una especificidad del 17 %[13]. Como puede verse, no hay relación entre sensibilidad y especificidad, así que una prueba puede tener las dos características elevadas, o ambas bajas, o alguna alta y la otra baja.

Existe, sin embargo, una situación en la que hay una relación absoluta de dependencia. En los casos en los que el resultado no es dicótomo *per se* y se toma un punto de corte para diferenciar al sujeto sano del enfermo (como ejemplo, el límite de 125 mg/dL en la glucosa para el diagnóstico de diabetes *mellitus*), la sensibilidad y la especificidad están claramente relacionadas, y la decisión del punto de corte es un compromiso.

Si el punto de corte fuese más bajo (por ejemplo, 100 mg/dL), se obtendría un examen con una mayor capacidad para identificar a los sujetos diabéticos como tales (mejora la sensibilidad), ya que aumentó la capacidad para identificar la enfermedad. Sin embargo, al haber reducido el límite aumenta la probabilidad de falsas clasificaciones (disminuye la especificidad), pues se diagnosticará como diabéticos a sujetos con valores ligeramente superiores a 100, que pueden corresponder a variaciones normales o a resistencia a la insulina sin diabetes.

Si el punto de corte fuese más alto (por ejemplo, 150 mg/dL), habría una mayor certeza de que no habrá diagnósticos falsos positivos (mejora la especificidad), pero se reduce la opción de diagnosticar a los sujetos diabéticos con valores en el extremo inferior (disminuye la sensibilidad).

Como puede verse, ***en todos los casos en*** *los que hay un punto de corte, la sensibilidad y la especificidad sí están íntima e inversamente relacionadas*: un cambio en el punto de corte siempre producirá una mejoría en una de las características a costa de la reducción en la otra. Mejora la sensibilidad, porque puede detectarse a un número mayor de sujetos enfermos, pero disminuye la especificidad porque se clasifica inadecuadamente a un número, también mayor, de sujetos sanos como enfermos.

CURVAS DE CARACTERÍSTICAS OPERATIVAS PARA EL RECEPTOR (CURVAS COR)

En los casos en los que puede decidirse un punto de corte, la decisión no debe ser arbitraria, y debería basarse en la selección del punto de corte con las mejores características: tanto sensibilidad como especificidad. Para determinar el mejor punto de corte, se calculan la sensibilidad y la especificidad de cada uno de los valores posibles, y se construye una gráfica en la que en el eje de ***y*** está la sensibilidad, y en el eje de ***x*** está la tasa de falsos positivos (1-especificidad). Puede verse en la **figura 10.6** la representación gráfica de las sensibilidades y las especificidades de los posibles puntos de corte de un examen.

A medida que aumenta la sensibilidad, disminuye la especificidad, y habrá un punto de corte donde la combinación ofrezca los mejores valores. Tradicionalmente, se lo describe como el punto más alto e izquierdo de la gráfica, donde se reúnen la mejor sensibilidad y la mejor especificidad.

Si una prueba no puede diferenciar al sujeto normal del anormal, equivale a los resultados que podrían obtenerse por azar. Así, se esperaría que la mitad de los controles tuvieran un resultado mayor que un paciente de quien se sabe que tiene la enfermedad, y la otra mitad tendrían un resultado menor. El área bajo la curva sería, entonces, de 0,5 (equivalente a la mitad de la gráfica, dividida por la diagonal a 45°), que es el menor valor posible.

Como una interpretación intuitiva del área bajo la curva, si el área es 1,0 puede decirse que todos los pacientes tendrán un resultado más anormal que los controles sanos, y si el área equivale a 0,8 puede decirse que, en promedio, un paciente tendrá un resultado más anormal que el 80 % de los controles.

En una curva de características operativas para el receptor se hacen evidentes los siguientes conceptos:

- A medida que aumenta la sensibilidad, disminuye la especificidad (y viceversa).
- Cuanto más cerca esté la curva del extremo superior izquierdo de la gráfica, mayor será la exactitud de la prueba.

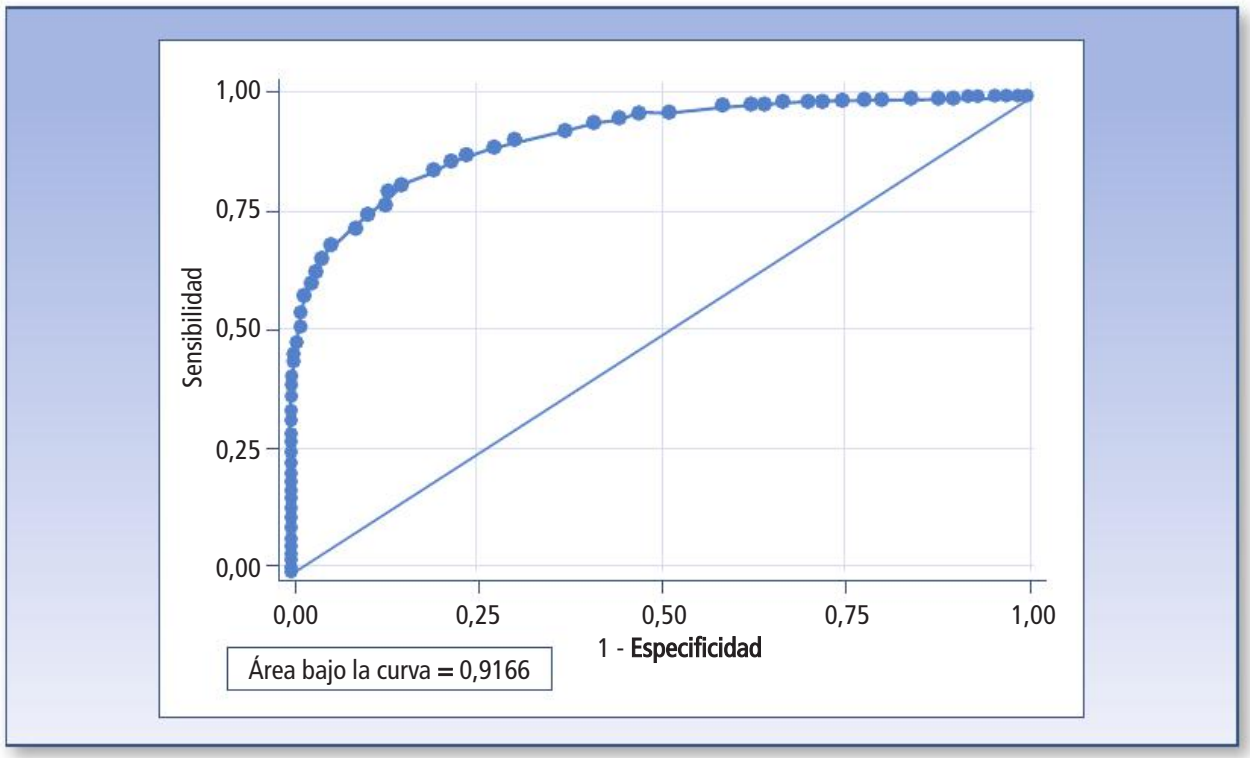


Figura 10.6 Curva de características operativas para el receptor (COR).

- Por lo anterior, cuanto mayor sea el área bajo la curva, mejor será la exactitud.
- Cuanto más cerca esté la curva de la diagonal a 45°, menor será la exactitud de la prueba.

Dado que la curva COR se utiliza no solamente para la evaluación global de la prueba (o para comparaciones entre pruebas), sino también para la decisión sobre el mejor punto de corte, podría pensarse que el mejor punto de corte es el que tenga la mejor sensibilidad y la mejor especificidad, y el cual, en la gráfica, corresponderá al punto más cercano al vértice superior izquierdo (el 100 % de sensibilidad y el 0 % de tasa de falsos positivos). Sin embargo, la decisión del punto de corte depende de cuán importantes sean la sensibilidad y la especificidad, o, en otros términos, de los riesgos que haya según la presencia de falsos negativos o de falsos positivos.

En algunas condiciones, como el diagnóstico de IAM en un servicio de urgencias, es, claramente, más riesgosa la presencia de falsos negativos que la de falsos positivos, de manera que el punto de corte ideal en una prueba de marcadores debería escoger la mejor sensibilidad posible que se acompañe de una especificidad apropiada. Claramente se debería privilegiar más la sensibilidad sobre la especificidad en este caso.

En la actualidad, hay programas de computador que permiten determinar los puntos de corte con la mejor combinación de características operativas, que facilitan la selección visual, pues presentan la tabla de sensibilidad y especificidad para cada punto de corte, y presentan la curva COR, además de calcular el área bajo la curva.

INTERPRETACIÓN DE RESULTADOS EN LA CLÍNICA

El paso anterior (determinación de la sensibilidad y la especificidad de una prueba) corresponde a la fase de investigación sobre el desarrollo de una prueba. Sin embargo, en la clínica el médico se enfrenta a una situación diferente. El paciente llega con una serie de manifestaciones y quejas que hacen sospechar una condición; el médico pide un examen y debe interpretar el resultado a la luz de la sospecha clínica previa a la solicitud de la prueba.

Si un clínico enfrenta a un paciente con muchos factores de riesgo para una enfermedad y pide un examen cuyo resultado es positivo, la interpretación de este resultado positivo es diferente de la interpretación del mismo resultado positivo obtenido en un sujeto sin los factores

de riesgo y con baja probabilidad de tener la condición. Es claro que la interpretación de un valor aislado de glucemia de 157 mg/dL no puede ser la misma en una persona obesa, de 59 años, con polidipsia y poliuria que en un joven de 27 años, con índice de masa corporal (IMC) normal y sin ningún antecedente personal o familiar de relevancia y en quien el hallazgo es parte de un examen de rutina. En el primer caso, la probabilidad de que el paciente tenga diabetes ya es alta antes del examen, y el examen parecería confirmar esa sospecha. En el segundo caso, sin embargo, la probabilidad es muy baja antes del examen, y la conducta del médico, probablemente, sea repetir el examen y asegurarse de que no hay un error.

A pesar de tratarse de dos resultados del mismo examen (ambos positivos), la interpretación diferente deja al descubierto la presencia de un tercer factor fundamental en la práctica clínica: la probabilidad de tener la enfermedad antes de hacer el examen, y que está determinada por la presencia de factores de riesgo, por los hallazgos clínicos o por determinantes genéticos o ambientales.

VALORES PREDICTIVOS

Cuando el médico recibe el resultado del examen, previamente debe haber determinado la probabilidad de que exista la enfermedad[14], bien sea a través de la prevalencia en la población, o bien, a través de los factores ya mencionados (riesgo, hallazgos físicos o en interrogatorio, determinantes genéticos o ambientales). Si se calcula que, dados los resultados del interrogatorio y de los factores de riesgo, la probabilidad preexamen del sujeto es del 80 %, ello equivale al concepto de ***frecuencia***, más fácil de entender por parte de los clínicos: si se tuviese una población de 100 sujetos iguales a nuestro paciente, 80 de ellos tendrían la enfermedad.

Sobre la base de este cálculo (o el uso de los datos de prevalencia para la población), se hace un cálculo que puede verse en la **figura 10.7**. Si se asume una población total de 1000 individuos (puede ser cualquier número), el 80 % de ellos tendrán la enfermedad (800 pacientes, ya que esa es la prevalencia o el cálculo de probabilidad preexamen). Si se decide la aplicación de una prueba que tenga una sensibilidad del 90 % y una especificidad del 95 %, los 800 sujetos enfermos serán clasificados correctamente en un porcentaje igual a la sensibilidad (90 %; habrá 720 verdaderos positivos y 80 falsos negativos). Y los 200 sujetos sanos serán clasificados, adecuadamente, en el 95 % (valor de la especificidad), con lo cual habrá 190 verdaderos negativos y 10 falsos positivos.

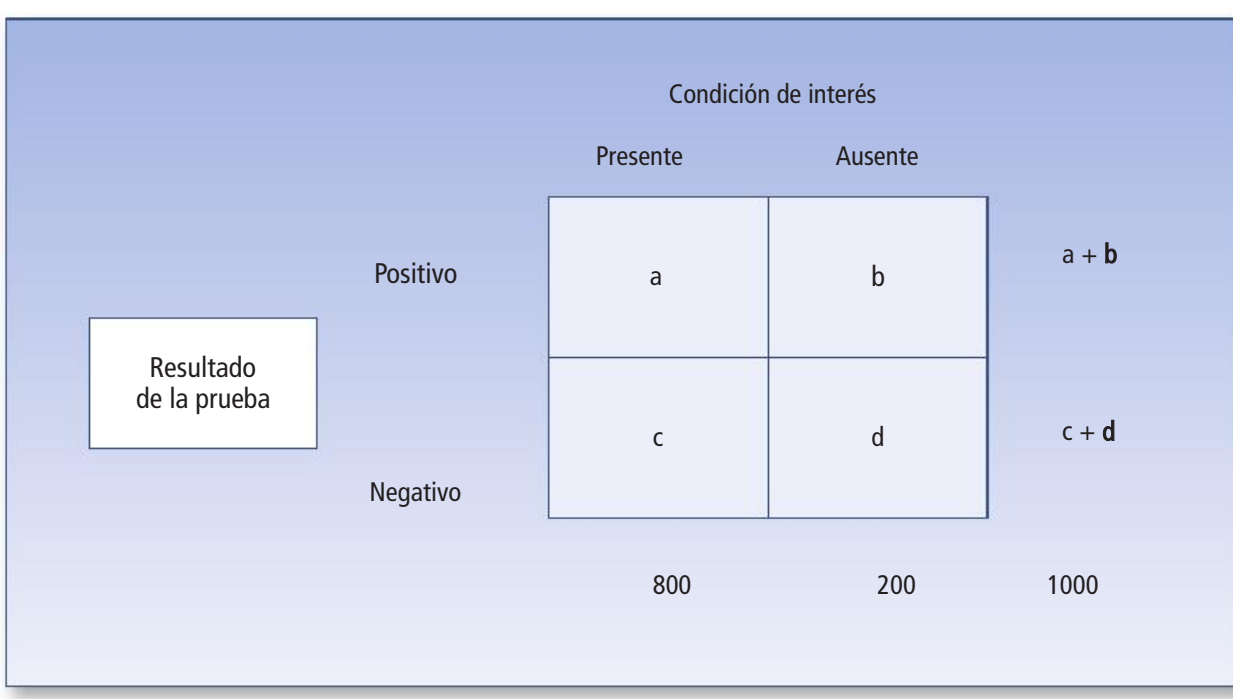


Figura 10.7 Cálculo de valores predictivos.

Valores predictivos de un resultado positivo

Con estos nuevos valores puede interpretarse el resultado del paciente. Para la interpretación de un resultado positivo, se utiliza solamente la información de los resultados positivos, como puede verse en la **figura 10.8**.

En la casilla ***a***, correspondiente a los verdaderos positivos, están los 720 sujetos correctamente clasificados (sensibilidad del 90 %) de los 800 que habría en la población ficticia de 1000, según la probabilidad establecida, del 80 %. En la casilla ***b*** están los 10 sujetos que tendrían resultado falso positivo entre los 200 sanos, según la especificidad del 95 %. Hay, en total, 730 resultados positivos, de los cuales 720 corresponden a resultados verdaderamente positivos, y 10, a errores en la clasificación.

Lo que le interesa al clínico del resultado positivo es determinar cuál es la probabilidad de que ese resultado positivo corresponda, en realidad, a personas con la enfermedad. Para determinar esto, se calcula el ***valor predictivo positivo de un resultado positivo***, obtenido a través del cálculo del porcentaje del total de resultados positivos (a + b) que corresponden a sujetos realmente enfermos (a). Este cálculo ***a / (a + b)*** ofrece la probabilidad de tener la enfermedad, dado que se tiene un resultado positivo, y da la respuesta a la inquietud del clínico. En el ejemplo, 720 / (720 + 10) = 98,6 %.

Hay dos posibles resultados: 1) el ***valor predictivo positivo de un resultado positivo***, que en el ejemplo se interpreta como que, ante un resultado positivo, en el paciente con el 80 % de probabilidad preexamen de tener la enfermedad, hay una probabilidad del 98,6 % de que el resultado positivo corresponda realmente a presencia de la enfermedad.

2) El otro posible resultado obtenido del análisis de los resultados positivos es el ***valor predictivo negativo de un resultado positivo***, que en el ejemplo es el complemento del anterior (100 – 98,6 = 1,4 %), que puede calcularse también con la fórmula ***b / (a + b)***. La interpretación es que ante un resultado positivo, la probabilidad de que no corresponda con la enfermedad es del 1,4 %. O, expresado de otra manera, de cada 100 pacientes iguales al anterior, que tengan un resultado positivo, solamente el 1,4 % corresponden a un resultado falso positivo.

Valores predictivos de un resultado negativo

Si el resultado del examen es negativo, también el clínico tiene la pregunta de si el resultado negativo corresponde, en realidad a ausencia de la enfermedad o no. Los resultados, de acuerdo con la probabilidad preexamen del ejemplo, se muestran en la **figura 10.9**.

Si el resultado es negativo, del total de resultados negativos (270), 190 corresponden a sujetos que verdaderamente están libres de la enfermedad estudiada, pero 80 tienen un resultado negativo, a pesar de estar enfermos. Si calculamos el porcentaje de negativos que corresponden a verdaderos negativos, d / (d + c) obtendremos un valor 190 / (190 + 80 = 70,3 %), que representa la probabilidad de que el resultado negativo corresponda realmente a ausencia de la enfermedad. Es el ***valor predictivo negativo de un resultado negativo***. En el caso del ejemplo, el resultado del 70,3 % significa que, ante un resultado negativo, en el sujeto con probabilidad preexamen del 80 %, la probabilidad de que el resultado corresponda real-

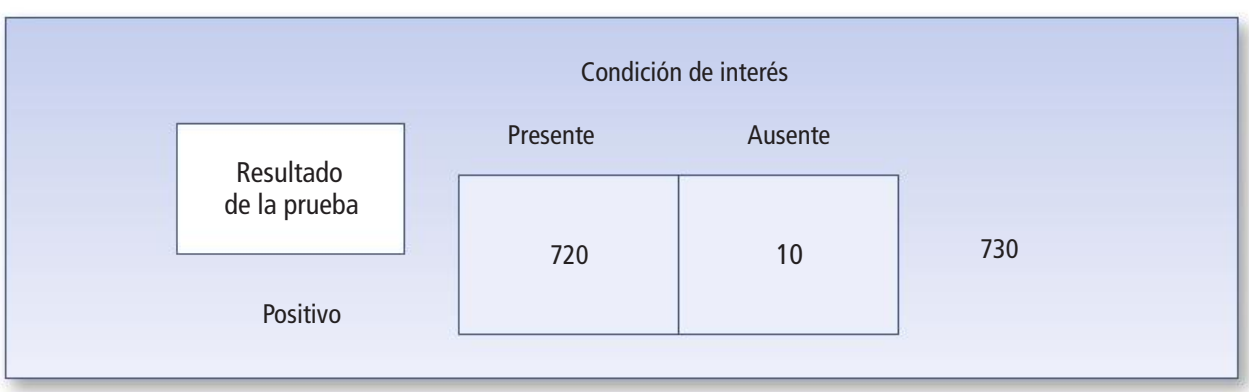


Figura 10.8 Cálculo de valores predictivos de un resultado positivo.

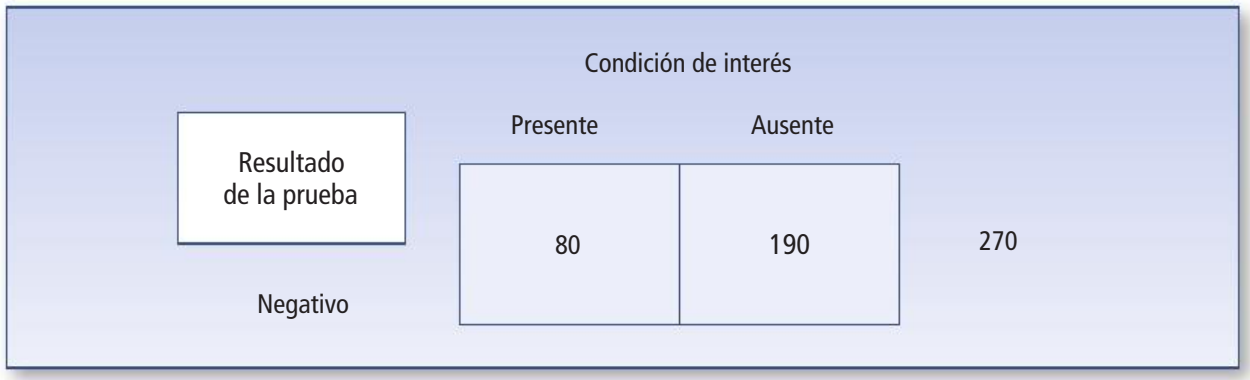


Figura 10.9 Cálculo de valores predictivos de un resultado negativo.

mente a ausencia de enfermedad es del 70,3 %. Los mismos datos permiten concluir que ante un resultado negativo, aún existe una probabilidad del 29,3 % (100 – 70,3) de que el sujeto tenga la enfermedad, que es el ***valor predictivo positivo*** *de un resultado negativo*.

Como puede verse, interpretar los valores predictivos negativos y positivos le permite al clínico tomar las decisiones adecuadas respecto a la interpretación de los resultados. Sin embargo, a pesar de la utilidad de los valores predictivos, estos no se encuentran en la literatura como sí se encuentran la sensibilidad y la especificidad de los exámenes. La razón está en que los valores predictivos dependen de la prevalencia o de la probabilidad preexamen, y no son extrapolables a todas las situaciones, como sí lo son, en general, la sensibilidad y la especificidad. Para cada valor de probabilidad preexamen se pueden y se deben calcular los valores predictivos positivos y negativos.

RAZONES DE PROBABILIDAD

Los conceptos evaluados hasta ahora (sensibilidad, especificidad y valores predictivos) tienen aplicación en los casos en los que puede hacerse la clasificación dicotómica de los resultados, como positivo y negativo, o presente y ausente. Sin embargo, con frecuencia se tienen resultados más complejos en medicina. Ya se vio cómo los resultados continuos (glucemia, por ejemplo) pueden convertirse en resultados dicotómicos a través de la selección de un punto de corte. Sin embargo, no pocas veces se pierden datos en esta transformación, y podría quererse la opción de tener una clasificación intermedia de severidad.

Una forma adicional de considerar la probabilidad de presencia de la enfermedad es a través del ***cálculo de las suertes***, que son la razón de dos probabilidades. Puede hacerse interconversión entre suertes (los ***odds*** de la literatura anglosajona) y probabilidades, así:

$$\text{Suertes} = \frac{\text{Probabilidad}}{1 - \text{Probabilidad}}$$

$$\text{Probabilidad} = \frac{\text{Suertes}}{1 + \text{Suertes}}$$

La razón de probabilidades se calcula para cada valor particular de una prueba no dicotómica, y corresponde a la división de la probabilidad del resultado de una prueba en presencia de la enfermedad por la probabilidad del resultado en personas sin la enfermedad.

$$\text{Razón de probabilidades} = \frac{\text{Probabilidad de un resultado en enfermos}}{\text{Probabilidad del mismo resultado en sanos}}$$

Las razones de probabilidades pueden calcularse para el resultado positivo, o para el resultado negativo; también, para resultados intermedios (si hubiera una categoría de resultado dudoso, la razón de probabilidades sería la probabilidad de tener resultado dudoso en quienes tienen la enfermedad dividida por la probabilidad del resultado dudoso e<-1<11<n quienes no tienen la enfermedad.

La interpretación final permite saber cuántas veces más (o cuántas veces menos) probable es que se pueda encontrar un resultado determinado de una prueba en personas enfermas, en comparación con que se le encuentre en personas sin la enfermedad. Una razón de probabilidades de 2,8

se interpreta como que es 2,8 veces más probable que resulte positiva una prueba en presencia de la enfermedad que en su ausencia.

La **tabla 10.1** muestra un ejemplo del uso de la gammagrafía de ventilación-perfusión en el diagnóstico de tromboembolismo pulmonar (TEP). Los resultados de la prueba se informan en cuatro categorías ordinales, como *alta probabilidad, probabilidad intermedia, baja probabilidad y resultado normal.*

Los resultados de la gammagrafía de ventilación-perfusión pueden interpretarse usando la razón de probabilidades para cada resultado posible de la gammagrafía, utilizando los datos del estudio del grupo Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED)[15,16]. Los resultados pueden interpretarse de acuerdo con el resultado así: si se obtiene una gammagrafía de probabilidad alta, es 13,9 veces más común tener este resultado si se tiene un TEP que si no se lo tiene. Como puede verse, la interpretación es más completa, pues tiene en cuenta todas las opciones de respuesta del examen y puede considerarse más de las clásicas dos respuestas dicotómicas, positiva o negativa[17].

Razones de probabilidad positivas y negativas

Pueden definirse dos tipos de razones de probabilidad, según el resultado del examen que se tome para los cálculos. En la *razón de probabilidades positivas* se comparan la probabilidad de un resultado positivo en sujetos enfermos y la probabilidad de un resultado positivo en sujetos sanos. Y en la *razón de probabilidades negativas* se toman las probabilidades de un resultado negativo en enfermos y la probabilidad del mismo resultado en sanos.

Interpretación de las razones de probabilidad

Los resultados de las razones de probabilidad pueden clasificarse en tres categorías:

- Razón de probabilidades = 1,0: Significa que las probabilidades del resultado son iguales en quienes tienen y quienes no tienen la enfermedad, de manera que el resultado de la prueba no va a ayudar en el diagnóstico, pues no cambia la probabilidad preexamen.
- Razón de probabilidades > 1,0: Este resultado aumenta la probabilidad de enfermedad, y se han propuesto límites, que se presentan en la **tabla 10.2**.
- Razón de probabilidades < 1,0: Este resultado disminuye la probabilidad de enfermedad, y también se han propuesto límites para su interpretación **(tabla 10.2)**.

Uso de las razones de probabilidad en medicina clínica

Las razones de probabilidad pueden ser muy útiles en el contexto clínico; particularmente, en pruebas diagnósticas con resultados no dicotómicos (aunque pueden utilizarse también en ese tipo de resultados). Como una ventaja de su uso, cabe mencionar que dependen menos de la prevalencia de la enfermedad o de la probabilidad preexamen que el enfoque con sensibilidad y especificidad.

Para su aplicación en la medicina clínica, se deben cumplir los pasos siguientes:

1. Cálculo de la probabilidad preexamen.
2. Conversión de la probabilidad preexamen en suertes preexamen.
3. Información sobre la sensibilidad y la especificidad de la prueba.
4. Cálculo de las razones de probabilidad utilizando la sensibilidad y la especificidad.
5. Cálculo de las suertes postexamen.
6. Conversión de las suertes postexamen en probabilidad postexamen.

Tabla 10.1 Razones de probabilidades: tromboembolismo pulmonar y gammagrafía de ventilación-perfusión (V/Q)-Estudio PIOPED

Gammagrafía V/Q	Presencia de TEP	Ausencia de TEP	Razón de probabilidades
Probabilidad alta	102	14	13,9
Probabilidad intermedia	105	217	0,93
Probabilidad baja	39	199	0,37
Normal	5	50	0,19
Total	251	480	

Tabla 10.2 Interpretación de las razones de probabilidades

Razón de probabilidades	Interpretación *Cambio en la probabilidad de la enfermedad*
< 10	Aumento significativo
5,0-9,99	Aumento moderado
2,0-4,99	Aumento pequeño
1,01-1,99	Aumento mínimo
1,0	No hay cambio
0,99-0,5	Reducción mínima
0,49-0,2	Reducción pequeña
0,19-0,1	Reducción moderada
< 0,1	Reducción significativa

Con un ejemplo clínico se desarrollará el proceso completo. En una mujer de 62 años de edad, hipertensa y con hipercolesterolemia en tratamiento, se presenta dolor torácico típico. Sin otros datos relevantes, se hace un cálculo de probabilidad preexamen de un IAM del 83,9 % utilizando la tabla de probabilidad de la Sociedad Europea de Cardiología[18]. Se utiliza un biomarcador para diagnóstico de IAM cuyas características operativas, informadas por el fabricante, son: del 92 % para sensibilidad y del 95 % para especificidad.

1. **Probabilidad preexamen:** Del 83,9 %, que puede expresarse también como 0,839.
2. **Conversión de la probabilidad preexamen a suertes preexamen:** 0,839 / (1 – 0,839) = 5,21
3. **Sensibilidad de la prueba: del 92 %; especificidad: del 95 %.**
4. **Cálculo de la razón de probabilidades (para un resultado positivo):** sensibilidad / (1 – especificidad) ≥ 0,92 / (1 – 0,95) = 18,4.
5. **Cálculo de las suertes postexamen** = razón de probabilidades × suertes preexamen 18,4 × 5,21.
6. **Conversión de suertes postexamen en probabilidad postexamen:** 95,864 / (1 + 95,864) = 0,98.
7. La probabilidad de infarto en la paciente, que antes de la prueba era del 83,9 %, se convierte, dado el resultado positivo, en una probabilidad postexamen de 98%.

PRUEBAS MÚLTIPLES

Puesto que las pruebas diagnósticas utilizadas no tienen siempre características operativas muy satisfactorias, se hace necesario, a menudo, hacer múltiples exámenes, cuyos resultados deben evaluarse en conjunto. La primera decisión que debe tomarse es si las pruebas múltiples se hacen simultáneamente (***pruebas en paralelo***) o si se hacen en secuencia, de acuerdo con los resultados (***pruebas en serie***).

En el paciente en quien se necesita una rápida evaluación (sala de urgencias, cuidado intensivo) se prefiere hacer pruebas en paralelo, así como en los casos donde haya dificultades logísticas para llevar a cabo varias visitas y exámenes secuenciales. La estrategia permite aumentar la sensibilidad y el valor predictivo negativo, pues hay una menor probabilidad de no diagnosticar la enfermedad. Dicha estrategia corresponde a la utilizada en los centros de referencia, lo cual, a su vez, explica el mayor número de diagnósticos hallados respecto a los centros primarios. Simultáneamente, aumenta el riesgo de falsos positivos; es decir, se reducen la especificidad y el valor predictivo positivo.

Por el contrario, cuando hay tiempo y recursos para hacer la evaluación pausadamente y no hay urgencia, la estrategia de pruebas en serie resulta más específica y aumenta el valor predictivo positivo, al tiempo que disminuyen la sensibilidad y el valor predictivo negativo. La secuencia permite evitar exámenes riesgosos o muy costosos mediante la exploración previa con pruebas más económicas o más seguras. Se corre el riesgo de no hacer el diagnóstico, pero se tiene mayor certeza sobre la presencia de la enfermedad cuando el resultado es positivo.

CONCLUSIONES

En la investigación sobre pruebas diagnósticas se evalúan la sensibilidad y la especificidad. Estas características operativas se logran luego de la aplicación de un patrón de oro a un grupo de sujetos, que son, entonces, clasificados como enfermos o sanos (o libres de la enfermedad de interés). A estos sujetos, ya clasificados, se les aplica el examen en estudio, y los resultados permiten evaluar cuántos de los enfermos son correctamente clasificados por la prueba (sensibilidad) y cuántos de los sujetos sanos son también correctamente catalogados como sanos (especificidad).

Cuanto mayor sea la sensibilidad de una prueba, más útil será en la aplicación posterior

a la clínica un resultado negativo; y cuanto más alta sea la especificidad, mayor será la utilidad de un resultado positivo.

La sensibilidad y la especificidad no están relacionadas, excepto cuando se calculan puntos de corte en resultados continuos. En este caso, cualquier alteración del punto de corte hará mover la sensibilidad en el sentido contrario al sentido en el que se mueva la especificidad.

En la aplicación en la clínica, el concepto fundamental es la probabilidad preexamen de que exista la enfermedad (*prevalencia*). Sobre este concepto se aplican la sensibilidad y la especificidad, y los resultados permiten calcular los valores predictivos positivo y negativo, que tienen aplicación directa en la medicina.

En los casos en los que haya resultados no dicotómicos, el uso de razones de probabilidad permite conocer cuántas veces es más probable que se obtenga el resultado evaluado en presencia de la enfermedad que en su ausencia.

Referencias

1. Riegelman RK. Studying a study and testing a test. Reading evidence-based health research. 6th ed. Boston: Lippincott, Williams & Wilkins; 2012.
2. Sandler G. The importance of the history in the medical clinic and the cost of unnecessary tests. Am Heart J. 1980;100(6 Pt 1):928-31.
3. Collins RD. Differential diagnosis in primary care. 5th. ed. Philadelphia: Lippincott Williams & Wilkins; 2012.
4. Moons KG, de Groot JA, Linnet K, et al. Quantifying the added value of a diagnostic test or marker. Clin Chem. 2012;58:1408-17.
5. Turnbull JM. Is listening to abdominal bruits useful in the evaluation of hypertension? JAMA. 1995;274:1299-301.
6. Noguchi T, Yoshiura T, Hiwatashi A, et al. CT and MRI findings of human herpesvirus 6-associated encephalopathy: comparison with findings of herpes simplex virus encephalitis. AJR Am J Roentgenol. 2010;194:754-60.
7. Lindsay P, Connor Gorber S, Joffres M, et al. Recommendations on screening for high blood pressure in Canadian adults. Can Fam Physician. 2013;59:927-33.
8. Saslow D, Solomon S, Lawson HW, et al. Screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin. 2012;62:147-72.
9. Wolf A, Wender RC, Etzoni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: Update 2010. Ca Cancer J Clin. 2010;60:70-98.
10. Maity S, Nandi S, Biswas S, et al. Performance and diagnostic usefulness of commercially available enzyme linked immunosorbent assay and rapid kits for detection of HIV, HBV and HCV in India. Virology J. 2012;9:290.
11. Hwang SH, Oh HB, Choi SE, et al. [Meta-analysis for the pooled sensitivity and specificity of hepatitis B surface antigen rapid tests]. Korean J Lab Med. 2008;28:160-8.
12. Sullivan W, Vlodaver N, Tuna N, et al. Correlation of electrocardiographic and pathologic findings in healed myocardial infarction. Am J Cardiol. 1978;42:724-32.
13. Davie AP, Francis CM, Caruana L, et al. Assessing diagnosis in heart failure: which features are any use? Q J Med. 1997;90:335-9.
14. Bryant GD, Norman GR. Expressions of probability: Heuristics and biases. Science. 1974;185:1124.
15. The PIOPED Investigators: Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). JAMA. 1990;263:2753.
16. Stein PD, Henry JW. Prevalence of acute pulmonary embolism among patients in a general hospital and at autopsy. Chest. 1995;108:978.
17. Leeper KV Jr, Popovich J Jr, Adams D, et al. Clinical manifestations of acute pulmonary embolism. Henry Ford Hospital experience, a five-year review. Henry Ford Hosp Med J. 1988;36:29.
18. Genders TS, Steyerberg EW, Alkadhi H, et al., for the CAD Consortium. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. Eur Heart J. 2011;32:1316-30.



SECCIÓN 3

Diseños de investigación clínica. Estudios primarios

CAPÍTULO 11 La propuesta de investigación
Ana María de la Hoz Bradford, Jenny Alexandra Severiche Báez

CAPÍTULO 12 El protocolo de investigación. Una introducción
Luis Fernando Jaramillo García

CAPÍTULO 13 Introducción a los diseños de investigación: selección del diseño apropiado
Álvaro J. Ruiz Morales, Carlos Gómez-Restrepo

CAPÍTULO 14 Estudios de corte transversal
Andrés Duarte Osorio

CAPÍTULO 15 El experimento clínico
Juan Manuel Lozano León, Juan Gabriel Ruiz Peláez

CAPÍTULO 16 Estudio de casos y controles
Carlos Gómez-Restrepo

CAPÍTULO 17 Estudios de cohorte analítica
Juan Gabriel Ruiz Peláez

CAPÍTULO 18 Estudios sobre correlación y concordancia
Luis Felipe Uriza Carrasco, María Nelcy Rodríguez Malagón

CAPÍTULO 19 Estudios cualitativos
María Beatriz Ospina García, Carlos Gómez-Restrepo

CAPÍTULO

11

La propuesta de investigación

Ana María de la Hoz Bradford • Jenny Alexandra Severiche Báez

INTRODUCCIÓN

Los estudios de investigación suelen originarse en preguntas que buscan información para mejorar la toma de decisiones en la atención en salud.

Una vez se identifica una pregunta válida, y antes de empezar el estudio, el investigador debe definir el problema específico que se enfrenta, así como refinar la pregunta original.

Ello permitirá determinar la necesidad real de hacer el estudio, y si es así, permitirá formular de manera clara y factible una pregunta específica y seleccionar los métodos apropiados para lograr una respuesta válida y confiable.

Es probable que durante estos procesos que requieren la búsqueda de información adicional y el planteamiento de objetivos concretos, la pregunta original sufra modificaciones por los datos encontrados en la literatura. Podría, incluso, concluirse que la pregunta no necesita más estudios, pues fue respondida en investigaciones anteriores, o que no amerita ser evaluada, pues la información encontrada hace que no sea ya relevante, o bien, porque se hace cuestionable la factibilidad del estudio por limitaciones logísticas, en recursos o en tiempo.

El proceso de evaluación y refinamiento de la pregunta se desarrolla elaborando los documentos que definen de manera preliminar la justificación, las condiciones y los procedimientos necesarios para conducir un estudio: la *antepropuesta* **y la** *propuesta de investigación*. La elaboración de dichos documentos pretende, en términos generales, conducir y registrar el proceso para definir por qué y cómo se responderá la pregunta planteada.

La elaboración de la antepropuesta, como su nombre lo indica, precede a la elaboración de la propuesta, y es el primer paso en el planteamiento de una investigación.

Una vez cuenta con una antepropuesta satisfactoria, el investigador procederá a elaborar la propuesta, la cual amplía la justificación y la metodología de investigación presentadas en la antepropuesta, y que será la base para desarrollar el protocolo de investigación; este, a su vez, describe

con el máximo detalle los aspectos metodológicos y logísticos para conducir el estudio (véase el capítulo 12 "Protocolo de investigación"). En la tabla 11.1 se describen las principales características de la antepropuesta, de la propuesta y del protocolo de investigación, información que será ampliada en las siguientes secciones.

Un proceso juicioso de elaboración de la antepropuesta, la propuesta y el protocolo de investigación garantizará que todos los elementos importantes para conducir un estudio se hallen definidos e incorporados de manera adecuada y rigurosa al proceso, y facilitará que se obtenga financiación para ejecutar exitosamente la investigación planteada.

No hay en la literatura científica lineamientos universales o exclusivos para elaborar la antepropuesta ni para la propuesta; se describen aquí los elementos fundamentales para ello, y se ofrecen herramientas para generar propuestas con alta probabilidad de recibir financiación.

ANTEPROPUESTA

La antepropuesta es el acercamiento inicial a la concreción del problema y de la pregunta de investigación. Puede utilizarse como apoyo para tomar la decisión sobre el emprendimiento de un proyecto mayor. Los objetivos que deben seguirse durante el desarrollo de la antepropuesta son:

- Explorar la relevancia del problema.
- Hacer un primer ejercicio de refinamiento de la pregunta de investigación.
- Definir de manera preliminar los objetivos.
- Escoger la estrategia más apropiada para responder la pregunta planteada.
- Explorar la factibilidad de la conducción del estudio, así como las posibles fuentes de financiación y las alianzas estratégicas.

Para lograr tales objetivos, se sugiere que la antepropuesta contenga, al menos, los siguientes elementos: título, pregunta de investigación, justificación, objetivos, planteamiento metodológico y potenciales fuentes de financiación. La tabla 11.2 describe el contenido esperado para cada uno de dichos elementos.

Para elaborar una antepropuesta, se considera conveniente una búsqueda de la literatura, para caracterizar el problema e identificar la importancia de su abordaje. Con esta información se podrán identificar, de manera exploratoria, los vacíos en el conocimiento científico, se refinará la pregunta y se elegirá la estrategia o el diseño de investigación más adecuado. Es altamente probable que identificar información adicional y desarrollar cada uno de los elementos de la antepropuesta generen la necesidad de modificar la pregunta de investigación, y, en consecuencia, los objetivos de la investigación.

La antepropuesta permitirá decidir si se continúa el proceso y si se pasa a la elaboración de la propuesta; también, si se corroboró la importancia del problema planteado, si se concluyó que hay factibilidad del proyecto y si hay interés suficiente por parte del equipo de investigación. O se podrá decidir también abandonar la propuesta inicial o modificarla cuando se encuentre que el problema no es relevante, que la pregunta ya ha sido respondida, o que el estudio no es factible por cuestiones logísticas o por una baja probabilidad de financiación.

Tabla 11.1 Documentos necesarios para el planteamiento adecuado de una investigación

Documento	Descripción
Antepropuesta	**Documento de acercamiento inicial al problema de investigación, con la formulación de una pregunta** y un análisis preliminar de factibilidad del proyecto. Es la fase inicial, que permite plasmar las ideas y revisar la información y la logística del caso antes de emprender la tarea de elaborar una propuesta.
Propuesta	Presenta la estructura básica de una investigación y sirve como insumo principal para construir el protocolo de un estudio. Comunica la importancia de la pregunta y describe la estrategia seleccionada para responderla. Generalmente, se presenta para evaluación con el fin de recibir apoyo académico, técnico, financiero o logístico.
Protocolo	Es el documento donde se definen y se describen, de manera detallada y sistemática, los criterios de admisión, los objetivos, el diseño, los procedimientos para conducir el estudio, el plan de análisis y el plan de difusión de resultados. Se trata de un mapa muy detallado, que guiará la ejecución de la investigación.

Fuente: Elaboración propia.

Tabla 11.2 Secciones que debe contener una antepropuesta de investigación

Sección	Descripción del contenido
Título	Debe expresar la pregunta de investigación planteada, ser conciso y claro.
Pregunta de investigación	Debe presentarse de manera operativa, para que así permita identificar a la población de interés y las variables de estudio.
Justificación	Breve descripción del problema que enmarca la pregunta, y de los vacíos en el conocimiento que justifican conducir una investigación para responder la pregunta planteada.
Objetivo	Resultados que se buscan conduciendo una investigación. Lograr el objetivo permitirá responder de manera directa la pregunta planteada.
Planteamiento metodológico	Descripción breve del diseño que se considera más adecuado para responder la pregunta, y su temporalidad. Deben describirse la población de interés, el escenario de la investigación y un esquema inicial para conducir la investigación.
Potenciales fuentes de financiación	**Identificación de entidades o convocatorias que el investigador considera que** podrían financiar el estudio, y las potenciales fuentes de alianza y apoyo académico, administrativo y logístico.

Fuente: Elaboración propia.

PROPUESTA

Una vez terminado el proceso de antepropuesta, se pasa a desarrollar la propuesta de investigación, ya con una pregunta de investigación factible, concreta y relevante, enmarcada dentro de un problema específico.

La propuesta es un documento que, como su nombre lo indica, propone la estructura básica de la investigación, y que servirá, en caso de considerarse exitosa, como insumo principal para construir el protocolo del estudio. Una de las finalidades de la propuesta es comunicar la importancia de la pregunta planteada y proponer la estrategia adecuada para llegar a su respuesta. Por lo general, se presenta para evaluación (por evaluadores externos, comités de ética, instituciones de apoyo o, en alianza, entidades financiadoras) para recibir apoyo académico, técnico, financiero o logístico. En este orden de ideas, los objetivos de elaborar la propuesta de investigación son:

- Formular la pregunta de investigación definitiva.
- Identificar, caracterizar y sintetizar el problema general que enmarca la pregunta.
- Justificar la necesidad de responder la pregunta mediante un estudio de investigación.
- Plantear los objetivos de la investigación.
- Seleccionar el diseño de investigación más adecuado para responder la pregunta.
- Identificar los aspectos operativos, logísticos, administrativos y financieros iniciales que se requerirán para la conducción del estudio.

La rigurosidad, la claridad y la exhaustividad con las que se describa cada uno de los mencionados componentes aportará, en gran medida, al éxito de una propuesta.

La presentación de propuestas incompletas o poco claras llevará a la necesidad de aclaraciones sobre los contenidos, o al rechazo por parte de los evaluadores (y, por ende, a no obtener financiación o aprobación por parte del comité de ética) y a demoras adicionales en el proceso; tales factores disminuyen la probabilidad de que una propuesta se traduzca en una investigación real.

No es raro que las propuestas de investigación se desarrollen dentro del marco de convocatorias específicas para la financiación de proyectos de investigación, emitidas por entidades públicas o privadas, que exigen lineamientos específicos y requisitos particulares en el contenido y en la forma, para su sometimiento. Es necesario que los grupos de investigación se adhieran a los requerimientos de cada convocatoria particular, para aumentar la probabilidad de éxito

del documento presentado. A continuación se describen los componentes mínimos que deben ser incluidos en una propuesta de investigación exitosa **(tabla 11.3)**.

Identificación

Incluye la fecha, el título de la investigación propuesta, el nombre del investigador principal y los de los coinvestigadores, el nombre de las organizaciones a la cual pertenecen los investigadores o que avalan la investigación, y de las organizaciones o las instituciones aliadas para el desarrollo de esta. Siempre deben proporcionarse datos de contacto (usualmente, los del investigador principal): ciudad, teléfono, dirección y correo electrónico.

Título

El título de una propuesta cobra vital importancia, dado que constituye el primer contacto del lector o el evaluador con el documento. Debe ser conciso, pero suficientemente informativo para permitirle al lector identificar la naturaleza de la investigación, el área metodológica, la población de interés y las variables de estudio; es decir, que sugiera en su contenido la pregunta de investigación. Un título adecuado tiene el potencial de aumentar el interés de los evaluadores y de diferentes lectores, y contribuirá en la consecución de apoyo y de recursos.

Resumen

Describe de manera concreta los aspectos principales de la propuesta: planteamiento del problema, estado del arte y justificación de la investigación, los objetivos planteados, la estrategia escogida, o diseño, y el plan de análisis. Adicionalmente, puede convenir que se incluyan en el resumen los resultados esperados y su efecto o su impacto potenciales.

Tabla 11.3 Secciones que debe contener una propuesta de investigación

Identificación
Título
Resumen
Antecedentes
Objetivos
Metodología
Consideraciones éticas
Resultados esperados
Grupo de investigación
Cronograma
Referencias
Anexos

Fuente: Elaboración propia.

Antecedentes

También llamada ***introducción*** en algunas fuentes, esta sección busca contextualizar al lector con el problema de investigación y transmitir de manera clara la necesidad de conducir el estudio. Debe incluir tres elementos principales: la descripción del problema, el estado del arte y la justificación de la investigación, que pueden ser presentados de manera conjunta, como parte de la introducción, o pueden ser organizados como subsecciones separadas.

- **Debe exponer de** manera clara el problema de salud o la condición que generaron la pregunta de investigación, y cuál es su importancia; es decir, cuál es la carga local, regional o internacional de la condición o del problema de salud sobre la población. Puede ser útil reportar la frecuencia con la cual se presenta la condición en la población de interés y los factores asociados a ello, y describir cómo la condición afecta a las poblaciones más vulnerables. Se espera, con esta información, que el lector logre una visión más amplia del conocimiento que se tiene sobre el problema, y sus implicaciones en los pacientes y en la población general. No se considera conveniente incluir descripciones exhaustivas de las bases biológicas de la enfermedad, ni cadenas causales remotas, ni mecanismos fisiopatológicos ni mecanismos de acción moleculares de los tratamientos existentes. Se sugiere claridad en el contenido, y que su extensión se limite a la necesaria para que la información incluida enmarque el problema, con el fin de que el lector comprenda el contexto de la pregunta de investigación y su importancia.
- **El estado del arte debe hacer un** recuento y una descripción de la investigación relacionada con la pregunta planteada; debe incluir los resultados de estudios previos, la calidad de dichos estudios y aclarar si sus resultados son concluyentes. Para elaborar esta sección se deben hacer búsquedas exhaustivas en la literatura, de estudios

publicados y no publicados (locales e internacionales), para conocer esfuerzos previos, cuál es el estado actual del conocimiento en el área, y cuáles son los vacíos existentes. Pueden ser necesarias revisiones sistemáticas de la literatura para tener el panorama completo; sin embargo, no siempre es viable su conducción por el tiempo o por los recursos disponibles para elaborar una propuesta. La identificación de revisiones sistemáticas en la literatura y el reporte de sus resultados son ideales, pues ello evita duplicar esfuerzos por parte del equipo de investigación y provee gran cantidad de información, sintetizada de manera adecuada. Es importante presentar únicamente información de literatura seleccionada que sea pertinente y de adecuada calidad metodológica. La ausencia de estudios de buena calidad o relacionados con la pregunta es un hallazgo de gran importancia, que debe reportarse y contribuye de forma importante a la justificación de la propuesta. Se debe describir, de manera concisa, la estrategia de búsqueda empleada para identificar los estudios y los resultados obtenidos. Para las revisiones sistemáticas de la literatura se debe presentar un informe completo de la metodología y de los resultados, en un documento anexo a la propuesta.

- Justificación de la investigación: **La conclusión de** la introducción o de la sección de antecedentes debe ser la justificación de la investigación. Una vez planteado el problema y confirmado que no ha sido resuelto, o que la respuesta existente no es satisfactoria, queda expresada la necesidad de la investigación planteada. En algunos casos puede considerarse necesario replicar un estudio por razones biológicas, culturales, de adherencia a las intervenciones, etc., que podrían explicar diferencias con los resultados hallados en poblaciones o contextos diferentes. Esos casos deben estar plenamente justificados, y las limitaciones en la aplicabilidad de la información existente deben ser suficientemente descritas.



Es importante mencionar el aporte esperado de la investigación al cuerpo del conocimiento científico del problema de salud, y cuál es el posible impacto sobre la práctica real.

Además de llenar vacíos en el conocimiento, y de responder a la pregunta planteada, la conducción de un estudio puede traer beneficios agregados, como: fortalecer la capacidad académica e investigativa de las instituciones y de los grupos de investigación; contribuir a la formación del personal de la entidad ejecutora y de entidades u organizaciones aliadas; desarrollar capacidades regionales o nacionales, y aportar a la construcción de un cuerpo de conocimiento en investigación en el país. De manera complementaria, puede ser importante resaltar la relación de la investigación con las prioridades en salud y los desarrollos regionales o nacionales, y el beneficio potencial para los usuarios de la información que se obtenga: profesionales de la salud, pacientes, población general o administradores en salud. Deben describirse los beneficios adicionales esperados para mejorar la probabilidad de financiación para la propuesta.

Objetivos

Son los resultados principales que se buscan conduciendo el estudio; la información que brinden permitirá responder la pregunta investigativa planteada y elaborar las conclusiones.

Los objetivos planteados en el estudio deben responder de forma directa la pregunta de investigación, y deben formularse en términos operativos (de la misma manera que la pregunta de investigación, como se describió en capítulos anteriores), con el fin de seleccionar la estrategia de investigación (diseño) más adecuada para alcanzarlos.

Generalmente, se plantean dos tipos de objetivos para un estudio:

- Objetivo general: **Resultado principal que espera** alcanzarse al conducir la investigación. Debe responder directamente la pregunta de investigación.
- Objetivos específicos: **Resultados particulares o** específicos que deben alcanzarse con el fin de lograr el objetivo general (normalmente, se plantea más de uno).

También pueden plantearse objetivos ***primarios* y *secundarios*. Un objetivo primario es** el que guía el diseño de la investigación, mientras que un objetivo secundario busca explorar información adicional dentro del marco de una investigación diseñada con una finalidad diferente. El planteamiento de objetivos secundarios

en una investigación es opcional, y dependerá de los intereses particulares del investigador o del grupo de investigación.

Metodología

Esta sección describe de manera general los procedimientos que se utilizarán para lograr los objetivos planteados. Deben abordarse los aspectos metodológicos y logísticos relacionados con el diseño escogido. A pesar de que la profundidad en la descripción metodológica incluida en la propuesta dependerá, en gran medida, de las características de cada convocatoria particular, en términos generales no suelen exponerse de manera exhaustiva ni muy detallada los pasos ni los procedimientos planeados para conducir la investigación, como sí deben plantearse en un protocolo.

Se recomienda que la propuesta incluya, al menos, la información necesaria que permita al evaluador juzgar si los métodos seleccionados para responder la pregunta son adecuados y factibles. Cada diseño de investigación requiere incluir elementos particulares en la metodología, que deben ser tenidos en cuenta. A continuación se describen los elementos que deben estar en una propuesta para para cualquier tipo de investigación, y los cuales se sugiere presentar en diferentes apartados, en aras de la claridad para el lector.

Diseño

En este apartado debe describirse con claridad el diseño seleccionado para responder la pregunta, y su temporalidad (de tipo ***histórico***, ***concurrente*** o ***mixto***). A partir de las características del diseño seleccionado, se estructurarán las descripciones de los siguientes elementos de la metodología.

Población

Deben definirse la población de interés del estudio y las características generales de la población a la cual se espera que los resultados del estudio sean generalizables. Es muy importante identificar cuál será la población sujeto de estudio, con los ***criterios de inclusión***; y cuáles, las personas que, no obstante cumplir con los criterios de inclusión, no podrán participar en el estudio, por diferentes razones (de seguridad, éticas, metodológicas o logísticas), por medio de los ***criterios de exclusión***.

El contexto, el lugar o los lugares donde se hará la convocatoria, el reclutamiento de pacientes y la estrategia para ello pueden ser descritos. Si es relevante, deben reportarse el lugar o los lugares donde se hará el muestreo. Estas descripciones ubicarán al lector del protocolo en el contexto en el cual la investigación será conducida.

Tamaño de la muestra

El tamaño de muestra estimado debe citarse, junto con el método utilizado para su cálculo, y aclarar si se hicieron ajustes por pérdidas potenciales de pacientes durante el seguimiento. Es ideal que se mencione la fuente informativa utilizada para el cálculo, y que la elección de dichos datos sea justificada.

Variables

La información sobre las variables indicará al lector qué se medirá durante el estudio y cómo se medirá. Cada una de las variables de interés para el estudio debe ser listada, y para cada una de ellas debe proveerse la definición operativa relacionada con su medición. Se debe describir cada variable según su naturaleza, según la manera como será medida (cualitativa o cuantitativa) y según la escala de medición (véase el **capítulo 4** "Variables y fuentes de error en la medición").

Es importante incluir la totalidad de las variables de interés: de exposición, de desenlace, variables demográficas, de confusión o de interacción. En esta sección debe incluirse la definición operativa de población expuesta, o la definición de un caso, así como la definición del control cuando se trate de estudios analíticos. Así mismo, deben describirse, en caso de que la investigación involucre intervenciones, cuáles son las características de la intervención en los grupos experimental y de control, la naturaleza de dicha exposición (farmacológica o no farmacológica, por ejemplo), la dosis, la frecuencia y las condiciones de administración, y si se utilizaron técnicas especiales, nuevas o convencionales.

Instrumentos

Todas las herramientas que servirán para medir variables y para recolectar los datos deben ser incluidas y descritas: cuestionarios, dispositivos, instrumentos de medición de variables biológi-

cas, escalas o pruebas diagnósticas, entre otros. Se deben describir las características de la medición, la validez y la precisión esperada de ellas cuando se utilizan los instrumentos propuestos. Si el estudio requiere herramientas que no son de libre uso, debe hacerse explícito que se cuenta con autorización o se dispone de los recursos para pagar los derechos correspondientes.

Procedimientos

Describe la totalidad de los pasos que deben cumplirse para conducir adecuadamente el estudio: desde la selección de los pacientes hasta la finalización del seguimiento y el análisis de datos. Debe incluir y describir con claridad cada uno de los procesos, en una secuencia ordenada y lógica para alcanzar los objetivos propuestos. Se sugiere incluir flujogramas que integren la totalidad de los procedimientos. Los procesos que normalmente están incluidos en este apartado son:

- Identificación y selección de los pacientes, o ***estrategia de muestreo***.
- Asignación o clasificación de los pacientes en estudios analíticos.
- Administración de la intervención.
- Mediciones (que pueden darse una sola vez o varias veces durante el estudio).
- Estrategia de seguimiento, visitas de evaluación o de intervención.
- Criterios para el retiro de un paciente.

Además de lo anterior, debe quedar claro para el lector cuándo y dónde se llevará a cabo cada uno de los procedimientos para generar la información requerida (mediciones) y cómo será recopilada, si se utilizarán formatos electrónicos o físicos, cómo se estandarizará la recolección de la información, cuándo será digitada, si la digitación contará con procedimientos para verificar su calidad y como se construirá la base de datos.

Adicionalmente, debe reportarse si el personal encargado de recolectar la información cuenta con entrenamiento previo o lo recibirá, si las mediciones las realizará una sola persona o varias, si los evaluadores fueron entrenados para estandarizar la medición y si las herramientas fueron probadas mediante estudios piloto para verificar su idoneidad. Si se planea utilizar algún tipo de programa informático para recolectar, procesar o consolidar la información, se recomienda incluir el nombre de dicho programa.

Por otra parte, se debe considerar la incorporación de un organigrama sobre la composición y el funcionamiento del grupo de investigación, así como la descripción del conducto establecido para la transferencia de la información, y el papel específico de cada uno de los investigadores y de los demás miembros del equipo mientras se conduzca la investigación. No es raro, principalmente en estudios de gran tamaño, tener manuales de procedimiento para el personal; en ese caso, se recomienda adjuntar el manual del caso al protocolo, en la sección de anexos.

Plan de análisis

El plan de análisis debe diseñarse y reportarse en función de los objetivos propuestos. La estrategia para el procesamiento de la información, la codificación y las medidas estadísticas deben definirse antes del inicio del estudio, así como los métodos estadísticos que se utilizarán. Es aconsejable incluir el nombre del programa que se utilizará para consolidar la base de datos y para el análisis.

Consideraciones éticas

Debe describirse el tipo de riesgo que pueden correr los participantes (sin riesgo, con riesgo mínimo, o con riesgo mayor que el mínimo), y qué elementos garantizarán que la conducción del estudio respetará los principios éticos promulgados por la Declaración de Helsinki de la Asamblea Médica Mundial[1] y la normatividad local o nacional para la investigación en salud[2]. También debe presentarse en detalle el proceso para el consentimiento informado e incluir el formato diseñado para tal fin, en caso de que vaya a obtenerse de manera escrita (como documento anexo a la propuesta). En caso de que los investigadores consideren que no es necesario obtener el consentimiento informado, o que este será obtenido por medio

[1] Colombia, Ministerio de Salud. Resolución 8430 de 1993, por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Bogotá: Minsalud; 1993.

[2] WMA. Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects [internet]. 2013 [citado 2014 ago. 22]. Disponible en: http://www.wma.net/en/30publications/10policies/b3/

de representantes de la comunidad, de organizaciones o de grupos poblacionales, dicha decisión deberá justificarse.

Los investigadores deben declarar y reportar sus intereses, para dar transparencia al proceso y permitirles a los evaluadores conocer un potencial conflicto de algún investigador (o del equipo de investigación) con el interés primario de la investigación que se propone.

Resultados esperados



Como resultado de una investigación se debe esperar la generación de conocimiento nuevo que responda la pregunta, de productos desarrollados desde los procedimientos de la investigación o de impactos potenciales que se darán a partir de la aplicación del conocimiento generado.

Es recomendable que la descripción de los resultados esperados se estructure de manera clara y en los términos anteriormente descritos, para contribuir a justificar la necesidad y la importancia de la investigación.

El nuevo conocimiento se obtendrá en función de los objetivos planteados, y puede ser descrito de esa manera; deben reportarse la información que será recolectada y la medida de resumen, o ***estimador***, que se pretende obtener para cada objetivo. Podría incluirse también una descripción de la ***generalizabilidad*** esperada para la información.

Un producto de la investigación puede ser un desarrollo propio, establecido desde el principio como objetivo de la investigación (una nueva tecnología, un dispositivo, una metodología, una herramienta o un instrumento diagnóstico o de medición, entre otros), o un vehículo para transmitir y difundir el nuevo conocimiento (publicaciones y eventos científicos para la difusión de los resultados, manuales, informes técnicos). Adicionalmente, puede esperarse que se generen productos relacionados con la investigación en sí misma, y no únicamente con sus resultados, como la capacitación del personal involucrado en la conducción del estudio, entrenamiento específico en investigación de estudiantes de pregrado y de posgrado y el fortalecimiento de la capacidad investigativa de las instituciones participantes.

La identificación de los impactos potenciales permitirá al lector tener una idea sobre el posible efecto del nuevo conocimiento en la toma de decisiones en la práctica clínica, en la calidad de la atención en salud o en la disponibilidad y la accesibilidad a los servicios, y, en última instancia, en el mejoramiento de la salud de los pacientes y de la población general. El impacto esperado justificará, en gran medida, la inversión de recursos en la investigación planteada.

Grupo de investigación

La composición del equipo investigador es un elemento de gran interés para los evaluadores externos, en cuanto a decidir sobre la asignación de recursos para un estudio. Si la propuesta es presentada por un grupo de investigación ya conformado, debe describirse de manera completa su trayectoria en investigación y la experiencia particular con la que cuenta en el área de interés para la propuesta presentada, además de su calificación.

Debe incluirse el listado de personas que conforman el grupo, junto con la descripción del papel que cumplirá cada uno en la conducción de la investigación. Debe incluirse, además, el listado de cargos adicionales con los que será necesario contar una vez se disponga de recursos para iniciar el estudio. A continuación se describen los papeles convencionales que comúnmente se encuentran listados dentro de un equipo de investigación:

- Investigador principal: Es el líder del equipo de investigación, y, generalmente, el autor principal de la propuesta. Suele ser experto en el área temática de la investigación y en la metodología propuesta.
- Coinvestigadores: Investigadores y expertos en el área temática, que aportan significativamente en la elaboración de la propuesta y participarán de manera activa en la mayor parte de las etapas de la investigación. Suele contarse con un número de entre dos y cinco coinvestigadores, aunque no hay al respecto un límite determinado que pueda plantearse *a priori*, y fijarlo dependerá del tamaño del estudio y del área de investigación.
- Coordinador: Investigador que se encarga de manejar y organizar aspectos logísticos, de distribuir periódicamente actividades, y de manejar y solucionar dificultades. Debe tener comunicación directa con el investigador principal, para transmitirle de

manera oportuna los avances, los informes y los problemas, para que así se puedan poner en marcha las acciones necesarias y se evite poner en riesgo la validez de la investigación.

- Asistente de investigación: Generalmente, un profesional de la salud con entrenamiento o experiencia en investigación, investigador júnior o estudiante. Se encarga de labores logísticas generales, de coordinar actividades durante la conducción del estudio, de recopilar y manejar la información, y de elaborar informes técnicos, entre otros. Además, trabaja de manera articulada con el coordinador y bajo la supervisión directa de este.
- Expertos en estadística y otras metodologías específicas: Participan en la investigación con aportes específicos durante las etapas que se requieran; por ejemplo, calculando el tamaño de la muestra, eligiendo la estrategia de muestreo o haciendo búsquedas de la literatura o análisis de datos.
- Asistente de administración: Es la persona encargada del manejo financiero de la investigación, de hacer contratos, compras y pagos.

Puede ser necesario incluir el currículo de los investigadores como anexo de la propuesta de investigación. Evaluar la propuesta suele incluir la verificación del conocimiento y la experiencia de los investigadores que la presentan; principalmente, en investigación relacionada con el área cubierta en la propuesta. Cada uno de los currículos debe incluir: títulos académicos, filiación o filiaciones institucionales y cargos que desempeña, así como la descripción de la experiencia previa en investigación, reflejada en trabajos y publicaciones previas en el área de interés.

Cronograma

El cronograma es un elemento importante a la hora de formular una propuesta de investigación; debe contener cada una de las actividades planeadas y cuándo serán llevadas a cabo durante el tiempo que dure la ejecución del estudio. Cualquier cambio, retraso o modificación en una o más de las actividades planeadas traerá consigo problemas en el desarrollo la investigación; por tal motivo, el cronograma debe elaborarse de manera cuidadosa, anticipando posibles problemas o retrasos a la hora de definir los tiempos de ejecución de cada actividad y de la investigación, de manera global.

El cronograma traduce la metodología del estudio a una lista de pasos y procesos que se deben cumplir para lograr los objetivos propuestos, y que describen el desarrollo del proyecto en un orden lógico.

A cada una de las actividades descritas se le debe asignar un periodo específico para su ejecución.

Usualmente, el cronograma se presenta como una tabla de dos ejes: el primero, con cada una de las tareas o actividades que se van a realizar, y el segundo, con el tiempo que tomará cada una de estas en la unidad de medida seleccionada **(figura 11.1)**. Es importante tomar en cuenta el tiempo total estimado o propuesto para la ejecución del proyecto, con el fin de determinar cuál será la unidad de medida utilizada a la hora de elaborar el cronograma. Para dicha tarea se pueden utilizar semestres, trimestres meses, semanas o días, dependiendo del detalle exigido o necesario en el cronograma. De igual forma, es importante tener en cuenta una fecha de inicio y una de finalización de los procesos, para tener un control cercano de cada una de las actividades propuestas y estimar, en términos globales, el impacto de los retrasos individuales en el desarrollo del proyecto.

Descripción: Ejemplo de figura para el cronograma de un estudio de investigación utilizando un diagrama de Gantt. La primera parte del diagrama (columnas de la izquierda) lista y numera en orden secuencial cada una de las actividades contempladas para ejecutar la investigación. La segunda parte (columnas de la derecha) se encuentra dividida en meses, y contiene barras horizontales de colores que indican el tiempo estimado de dedicación para cada actividad contemplada.

La herramienta más utilizada para elaborar los cronogramas es la hoja de cálculo de Microsoft Excel®, que permite presentar el cronograma como un *diagrama de Gantt* (diagrama de barras) o de *Pert* (diagrama de flechas). Existen también programas como Project®, que permiten un control y un seguimiento más detallados de cada una de las tareas, y la posibilidad de hacer un seguimiento de las actividades con asignación

Cronograma general de actividades

No.	Actividades a realizar: Descripción	Semanas	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
			1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
1	Revisión y ajuste del protocolo	8												
2	Diseño estadístico	6												
3	Selección de la muestra	2												
4	Diseño plan operativo	6												
5	Diseño instrumentos	6												
6	**Desarrollo del** *software* de captura y **validación de datos**	**6**												
7	Desarrollo *software* monitoreo y control	6												
8	Aplicación prueba piloto	1												
9	Evaluación prueba piloto y ajuste de instrumentos	1												
10	Capacitación nivel 1	2												
11	Capacitación nivel 2	2												
12	Vinculación del personal de campo	1												
13	Levantamiento, recuento y encuesta	9												
14	Seguimiento, monitoreo y control	11												
15	Consolidación, validación y depuración final de datos	6												
16	Ajuste factores de expansión	2												
17	Generación de archivos para análisis de resultados	4												
18	Análisis de resultados	13												
19	Elaboración documento final y entregables	4												

Figura 11.1 Ejemplo de un cronograma (diagrama de Gantt).

Fuente: Elaboración propia.

de recursos mucho más versátil que otro tipo de hojas de cálculo más sencillas.

Al elaborar el cronograma, debe tenerse en cuenta lo siguiente:

- Cada una de las actividades debe aparecer en orden cronológico, y debe quedar claro cuáles tareas son predecesoras de otras.
- Deben tenerse en cuenta posibles retrasos en la programación.
- Deben establecerse los recursos necesarios para cada actividad. Eso incluye una lista de cada actividad, con el responsable de su desarrollo y un listado de los demás recursos necesarios.
- Deben tenerse en cuenta los posibles ***tiempos muertos*** (vacaciones, fines de semana, temporadas, entre otros).
- En ocasiones, resulta de gran ayuda conocer los tiempos ejecutados en anteriores proyectos similares.
- Debe establecerse un margen de tiempo, en vez de periodos fijos e inflexibles; puede ser riesgoso establecer cronogramas muy justos, que no permitan ciertos eventos inesperados.

La elaboración del cronograma es una tarea que requiere precisión y constituye un elemento de vital importancia en la formulación de una propuesta, pues, por ser el insumo principal para elaborar el presupuesto, permitirá garantizar la disponibilidad de los recursos necesarios para la ejecución de cada actividad. Deben considerarse todas las variables, las actividades y los procesos indispensables para una correcta ejecución del proyecto, y debe buscarse toda la flexibilidad posible en el momento de su elaboración.

Presupuesto

Para garantizar la ejecución y la sostenibilidad de un proyecto, es primordial planificar el presupuesto, un elemento fundamental para formular una propuesta de investigación. El presupuesto presenta el monto total del dinero necesario para conducir el proyecto; y en caso de que los investigadores desarrollen la propuesta para buscar financiación externa, el monto solicitado a la entidad financiadora y el aporte, o contrapartida, que aportarán los investigadores o la entidad solicitante de los recursos.

El presupuesto se compone de una serie de gastos e ingresos. Respecto a los ingresos (principalmente, cuando se está desarrollando una investigación con recursos que provienen de fuentes externas) es fundamental conocer el plan de desembolsos de la entidad financiadora, para establecer el flujo de caja que permitirá planear y asegurar la totalidad de los gastos a lo largo de un periodo, lo que garantizará la viabilidad de los procesos y la ejecución del estudio.

La asignación de recursos a cada una de las actividades debe ser congruente con los compromisos y con los productos esperados y entregables derivados de la investigación. Para calcular el presupuesto de un proyecto, es importante tomar en cuenta todos los gastos en los que se incurrirá durante su ejecución, e incorporar los costos tanto ***directos*** como ***indirectos***; los primeros son recursos que se invertirán directamente en actividades propias del desarrollo de la propuesta, y los costos indirectos, los que se relacionan con las actividades realizadas en la institución ejecutora en torno a la gestión del proyecto. Como se ve, deben incluirse en el presupuesto los costos asociados a las actividades de gestión y apoyo administrativo propias del desarrollo de una investigación.

El presupuesto puede ser formulado por el investigador principal, por un coinvestigador o por personas de apoyo que conozcan en profundidad el proyecto. Esta actividad puede ser asesorada por una persona con conocimientos en el área financiera. Es importante que el investigador principal revise y evalúe el presupuesto, para evitar que se omitan detalles, actividades, gastos o rubros que podrían poner en riesgo la ejecución de la investigación.

Detalle del presupuesto

Es importante discriminar por rubros el presupuesto; esto permite planear, organizar y ejecutar de forma ordenada los recursos, y una rendición eficiente de cuentas. Todos los rubros que se anticipen como necesarios para conducir adecuadamente la investigación deben ser descritos y justificados en detalle, lo cual aumenta la probabilidad de financiación **(tabla 11.4)**.

Entre los rubros más importantes que componen un presupuesto se encuentran:

- Personal: Está conformado por el equipo desarrollador, que incluye a investigadores

Tabla 11.4 Ejemplo de un presupuesto discriminado por rubros específicos

Rubros	Financiado	Contrapartida		Total
		Ejecutoras	Otras	
Equipos	$207 000 000	$5 000 000	0	$212 000 000
Materiales	$132 500 000	0	0	$132 500 000
Personal científico	$634 060 000	$32 000 000	0	$666 060 000
Servicios técnicos	$1 152 920 000	0	0	$1 152 920 000
Viajes y gastos de viajes	$32 200 000	0	0	$32 200 000
Salidas de campo	$1 198 000 000	0	0	$1 198 000 000
Eventos académicos	$42 000 000	0	0	$42 000 000
Publicaciones y patentes	$51 000 000	0	0	$51 000 000
Programas informáticos	$46 200 000	0	0	$46 200 000
Costos operativos (hasta el 10 % del valor del proyecto)	$349 588 000	0	0	$349 588 000
Totales	**$3 845 468 000**	**$37 000 000**	**0**	**$3 882 468 000**

Fuente: Elaboración propia.

principales, coinvestigadores, expertos nacionales e internacionales, asistentes de investigación y auxiliares, entre otros, quienes participan directamente en el desarrollo del proyecto. En ocasiones se requiere el apoyo de asesores externos nacionales o internacionales. En este rubro se deben tener en cuenta las horas de dedicación de cada uno de los investigadores y el número de meses que durará su participación.

- Equipos: **Dispositivos necesarios para la** ejecución del proyecto. Estos pueden variar desde computadores hasta equipos de laboratorio sofisticados, dependiendo del tipo de investigación.
- Programas informáticos: **Licencias de programas** especializados que se requieran para el desarrollo de actividades particulares del proyecto.
- Materiales e insumos: **Elementos necesarios** para el desarrollo del proyecto, y que pueden incluir desde papelería e implementos de oficina hasta reactivos o insumos especiales.
- Viajes: **Gastos en los que se incurre por el** desplazamiento del personal del grupo de investigación para la desarrollar actividades propias del estudio. Pueden incluir gastos de transporte aéreo o terrestre en territorio nacional o internacional, así como los gastos de alojamiento y de manutención.
- Salidas de campo: **Gastos relacionados con los** traslados a zonas de intervención, o labores de campo distintas de las discriminadas en el rubro de viajes. Un ejemplo común de este tipo de gastos es la aplicación de encuestas en diferentes zonas del país.
- Servicios técnicos: **Gastos relacionados con la** prestación de servicios por parte de personas naturales o jurídicas, y que no hacen parte del equipo desarrollador contemplado en el rubro de personal. Ejemplos de este tipo de gastos son los relacionados con traducción, corrección de estilo, diagramación o asesorías externas, entre otros.
- Capacitación: **Gastos relacionados con el** entrenamiento de personal, si es requerido. Puede incluir capacitaciones al grupo de trabajo o la asistencia a simposios, congresos o entrenamientos particulares para uno o algunos de los investigadores.
- Bibliografía: **Gastos destinados a la compra** de artículos, libros, suscripción a bases de datos, o revistas que sean necesarios para ejecutar el proyecto.
- Publicaciones: **Gastos en los que se incurre** para difundir los resultados de la investigación. Puede incluir la publicación de artículos, libros, manuales, pósteres, entre otros.
- Costos operativos: **Gastos indirectos de administración**, relacionados con el desarrollo del proyecto, y que incluyen todo el apoyo administrativo y logístico para desarrollar la investigación.

No existe un formato único para elaborar los presupuestos ni para describir los rubros incluidos. Cada una de las entidades financiadoras solicita los presupuestos en formatos diferentes y con información diversa. Es de vital importancia que el grupo de investigación elabore su presupuesto teniendo en cuenta el formato específico que solicite la entidad financiadora a la que esté presentando la propuesta de investigación. Así mismo, se deben revisar las condiciones de las convocatorias o de la financiación de proyectos de cada entidad, en busca de los rubros no financiables, que varían ampliamente entre entidades, y deberán ser asumidos directamente por los investigadores o por la entidad ejecutora.

Herramientas para elaborar un presupuesto

Si bien algunas entidades financiadoras exigen la elaboración del presupuesto en ciertas aplicaciones propias, la hoja de cálculo de Excel® es la herramienta más utilizada para formularlo. Esta hoja de cálculo está disponible dentro de la ***suite*** básica de Microsoft Office®, y permite la introducción de valores y fórmulas prediseñadas. Mediante el uso de hojas de cálculo para elaborar un presupuesto, se pueden aprovechar funciones como las tablas dinámicas para presentar informes tanto numéricos como gráficos.

Lo primero por definir es el formato del presupuesto que se utilizará, y a partir de ese punto, una vez introducidos los valores de referencia, se pueden establecer varios escenarios posibles, que permiten tener una idea más amplia de los recursos necesarios y de las contingencias y los imprevistos que se pueden presentar a lo largo de la ejecución de un proyecto.

A la hora de establecer el formato de su presupuesto, tenga en cuenta, como primera medida, el formato solicitado por la entidad a la que se someterá la propuesta de investigación, y siga de manera estricta sus reglas y requisitos; refleje de manera transparente y justificada sus gastos estimados, con el fin de aumentar la probabilidad de tener éxito en la financiación de la propuesta.

En términos generales, la presentación de un presupuesto se hace de una manera global, alimentada por rubros específicos; se deben incluir los detalles para cada uno de los rubros en cuanto a definición, unidad, cantidad, y costo. La descripción detallada posibilitará un control y un seguimiento más seguros de los recursos antes, durante y después de ejecutarse la investigación.

Siempre que se elabore un presupuesto, tenga en cuenta:

- Se debe trabajar sobre hechos y cifras concretas, y, en lo posible, no sobre suposiciones.
- El cronograma y el presupuesto van de la mano; cada una de las actividades planeadas en el cronograma debe traducirse en cifras, y debe verificarse que todas las actividades incluidas estén, a su vez, consideradas dentro del presupuesto.
- Deben pedirse cotizaciones para incluir soportes y cifras exactas.
- Deben tenerse en cuenta los incrementos anuales del índice de precios de consumo (**IPC**), o inflación, que impactan de forma directa el presupuesto.
- Si se tienen cifras en monedas extranjeras, debe considerarse el comportamiento de la tasa de cambio; un cambio inesperado en ellas puede afectar el presupuesto.
- Se debe tener en cuenta el tipo de financiador de donde se obtienen los recursos, ya que cada uno de ellos tiene establecidas políticas propias para ejecutar recursos con fines de investigación.
- Detallar cada uno de los rubros permitirá tener un mejor control de cada uno de los aspectos económicos involucrados desde el principio.
- El seguimiento del presupuesto permitirá controlarlo mejor y prepararse para cualquier contingencia.
- No olvidar cuantificar en cifras los entregables y las obligaciones derivadas de la investigación.
- Recordar siempre establecer un flujo de gastos basado en el flujo de ingresos.

Referencias

La información consultada, y que sirvió como base para la construcción del documento, debe estar adecuadamente referenciada; el formato para la referenciación dependerá de los requisitos de la organización o convocatoria a la cual se presenta la investigación.

Anexos

Los elementos que deben anexarse a la propuesta son, por lo general, documentos y herramientas de soporte, los cuales podrán ser consultados por los evaluadores en caso de que haya interés o necesidad de tener información adicional a la presentada en el cuerpo del documento.

Algunos de los elementos que suelen encontrarse en esta sección son: cartas de aval institucional de la entidad ejecutora a la que pertenecen los investigadores, y de las entidades aliadas o en las que se planea conducir el estudio; certificaciones de aprobación del estudio por parte del comité de ética institucional; currículos del investigador principal y de los coinvestigadores, y figuras, flujogramas y cuadros de soporte que por las características del formato exigido para la propuesta no puedan incluirse en el cuerpo del documento (diagrama de Gantt, organigramas del equipo de investigación, esquemas de procedimientos), manuales de entrenamiento para el personal, formatos de recolección de datos, escalas o instrumentos de medición, formato de consentimiento y asentimiento informado, manuales de procedimientos, entre otros.

PROPUESTA DE INVESTIGACIÓN EXITOSA

Una vez se dispone de una propuesta exitosa y se considera que el estudio es no solo factible, sino viable, dadas las condiciones logísticas y de financiación ya alcanzadas, el grupo de investigación pasará al desarrollo del protocolo, en el cual se reflejará la planeación detallada de la metodología, los procedimientos y las condiciones que guiarán la conducción de la investigación.

Para que una propuesta se traduzca en una investigación real, es necesario contar con una financiación adecuada. Para ello, los investigadores deben someter su propuesta a diferentes convocatorias y evaluaciones realizadas por entidades públicas o privadas. Puede presentarse una propuesta para una convocatoria particular en un área específica, o puede preparase una propuesta espontánea para buscar financiación por entidades o convocatorias relacionadas con el área de investigación. De cualquier forma, es indispensable, para obtener éxito en la obtención de recursos, que la propuesta cumpla con los requisitos establecidos por la organización financiadora, con el fin de que sea considerada para un proceso de evaluación.

Existen diferentes tipos de entidades financiadoras, que pueden reunirse en cinco grandes grupos: 1) entidades gubernamentales nacionales e internacionales, 2) industria de la salud (industria farmacéutica y de tecnologías de la salud), 3) entidades académicas, 4) organizaciones profesionales y 5) entidades de caridad. En la **tabla 11.5** se presenta un listado de organizaciones internacionales que financian proyectos de

Tabla 11.5 Entidades financiadoras y servicios de apoyo para financiar proyectos de investigación

Entidad	Dirección electrónica
Grand Challenges in Global Health	http://www.grandchallenges.org/Pages/Default.aspx
Bill & Melinda Gates Foundation	http://www.gatesfoundation.org/How-We-Work/General-Information/Grant-Opportunities
Comisión Europea, Servicio de Información Comunitario sobre Investigación y Desarrollo ( )	http://cordis.europa.eu/home_es.html
International Development Research Centre ( )	http://www.idrc.ca/EN/Funding/Pages/default.aspx
European Research Council	http://erc.europa.eu/funding-and-grants
Unión de universidades de América Latina y el Caribe ( )	http://www.udual.org/index.html
International Foundation for Science ( )	http://www.ifs.se
Grants.gov	http://www.grants.gov
National Science Foundation ( )	http://www.nsf.gov/funding/
Special Programme for Research and Training in Tropical Diseases ( ) World Health Organization ( )	http://www.who.int/tdr/grants/en/
International Clinical Epidemiology Network ( )	http://inclentrust.org/inclen/funds.php

Fuente: Elaboración propia.

investigación en diferentes temáticas. Para cada región o país deben considerarse las entidades que apoyan investigación.

Obtener financiación para investigación no es una tarea fácil. La competencia entre grupos de investigación por una cantidad limitada de recursos hace que solo las propuestas de gran relevancia y alta calidad sean exitosas en tal sentido.

Cabe esperar que antes de tener éxito la propuesta sea rechazada una o más veces, lo cual obligará a los investigadores a revisar y reformular el documento, o, incluso, la pregunta, los objetivos y las metodologías planteados, con el fin de estructurar una propuesta que refleje de manera contundente el planteamiento de un proyecto de investigación necesario para responder una pregunta de investigación importante e innovadora, con un impacto potencial que justifique la inversión de los recursos solicitados. En la **tabla 11.6** se presenta un listado de consejos para preparar propuestas exitosas.

Tabla 11.6 Elementos clave para la elaboración y la presentación de propuestas exitosas

- Seguir de manera exacta las reglas de presentación de documentos a la convocatoria o a la organización, en cuanto a la forma y el contenido.
- Revisar propuestas previas que hayan tenido éxito en la misma convocatoria u organización a la que se está presentando; utilizarla como ejemplo para la construcción de la propia propuesta.
- Antes y durante la elaboración de la propuesta, contactar a la persona encargada del manejo de la convocatoria para resolver las dudas que surjan sobre la forma o los contenidos esperados en cada documento.
- Informarse sobre las características profesionales y laborales de los posibles evaluadores de la propuesta; construir los contenidos de una manera adecuada para el público esperado.
- Resaltar la importancia de la investigación; para ello, puede cuantificarse la magnitud del problema y describir el potencial impacto de los resultados de su investigación.
- Utilizar lenguaje sencillo en la redacción de la propuesta; enfocarse en los aspectos principales, ser conciso y expresar de manera clara los conceptos.
- Ser cuidadoso en la escritura de la propuesta; utilizar el lenguaje de manera adecuada, redactarla correctamente y evitar errores de tipografía y de ortografía.
- Asesorarse sobre los aspectos metodológicos principales, por medio de expertos que den retroalimentación sobre aspectos particulares (estadística, análisis de datos, diseño de investigación).
- Planificar con tiempo la elaboración de la propuesta, y así disponer de un periodo previo de revisión por los pares (colegas, profesores), para identificar errores y áreas por mejorar antes de someter la propuesta.
- Escribir un resumen completo, que transmita de manera clara cuál es el problema y su importancia, la necesidad de la investigación planteada, los objetivos, la metodología y el potencial impacto de los resultados.
- Evitar relegar a la sección de anexos información importante: es posible que dicha sección no sea revisada durante la evaluación de las propuestas.

Fuente: Elaborado a partir de Inouye y Fiellin, 2005.

Al-riyami A. How to prepare a research proposal. Oman Med J. 2008;23:66-9.

Berger AM, Moore TA. Effective grant writing. J Infus Nurs. 2011;34:167-71.

Conn VS. Welcome to the dark side of grant writing. West J Nurs Res. 2013;35:967-9.

Devine EB. The art of obtaining grants. Am J Health Pharm. 2009;66:580-7.

Fathalla MF, Fathalla MM. A practical guide for health researchers. Geneva: World Health Organization, Regional Office for the Eastern Mediterranean; 2004.

Higdon J, Topp R. How to develop a budget for a research proposal. West J Nurs Res. 2004;26:922-9.

Inouye SK, Fiellin D. An Evidence-Based Guide to writing grant proposals for clinical research. Ann Intern Med. 2005;142:274-82.

Klopper H. The qualitative research proposal. Curationis. 2008;31:62-72.

Koppelman GH, Holloway JW. Successful grant writing. Paediatr Respir Rev. 2012;13:63-6.

Merrill KC. Developing an effective quantitative research proposal. J Infus Nurs. 2011;34:181-6.

Organization World Health (WHO). Health research methodology: A guide for training in research methods. Geneva: World Health Organization; 2001.



CAPÍTULO

12

El protocolo de investigación. Una introducción

LUIS FERNANDO JARAMILLO GARCÍA

INTRODUCCIÓN

El diccionario de la Real Academia de la Lengua define el término ***protocolo***, entre otras acepciones, como plan escrito y detallado de un experimento científico, un ensayo clínico o una actuación médica.

En el lenguaje médico también recibe ese nombre el ***proyecto de investigación***, y es, en esencia, una guía que describe de forma adecuada y detallada el proceso de investigación que se tiene pensado realizar.

Un protocolo de investigación completo debe dar respuesta a las siguientes preguntas:

- ¿Qué se va investigar?
- ¿Quién lo hará?
- ¿Qué se sabe del tema?
- ¿Cuándo se hará?
- ¿Dónde se hará?
- ¿Por qué se debe hacer?
- ¿Para qué hacerlo?
- ¿Cómo se hará?
- ¿En cuánto tiempo se hará?
- ¿Cuánto costará?

Si la lectura del documento permite que estas preguntas sean contestadas de forma clara e inequívoca, puede tenerse la satisfacción de haber elaborado un buen protocolo de investigación, así como la seguridad de que se cuenta con el insumo básico para dar los pasos siguientes.

No solo no es posible ni viable emprender una investigación sin un adecuado protocolo (por la falta del plan de navegación necesario, porque no habrá una planeación de los aspectos logísticos, porque no habrá una definición precisa que permita procedimientos estandarizados, y por muchos otros aspectos). Hay otras dos razones muy importantes que hacen que el protocolo sea un requisito indispensable para empezar el procedimiento: no debe, por ninguna razón, iniciarse un proyecto investigativo sin la evaluación y la aprobación de un comité institucional de investigación y ética; además, será virtualmente imposible obtener financiación institucional o externa sin un protocolo sólido y bien fundamentado.

En consecuencia, la elaboración del protocolo siempre precede a cualquier trabajo de investigación y cumple tres funciones básicas: 1) comunicar los planes y las ideas del grupo o el líder de la investigación, 2) planificar al detalle todo el proceso y 3) comprometer al investigador, a su equipo y a la institución en la ejecución del proyecto.

Otra acertada concepción sobre la utilidad y la necesidad del protocolo de investigación es que este constituye la carta de navegación del

proyecto y contiene en detalle y profundidad todos los pasos que deben seguirse durante tal proyecto.

De lo dicho hasta el momento se deduce que un protocolo de investigación es un producto del conocimiento y de la reflexión sobre un tema o un problema objeto de estudio o interés, y en él está contenido lo que se quiere hacer, desde qué punto de vista se pretende hacer y cómo se hará, lo que constituye la esencia o el fondo del protocolo.

Ahora bien, así como el documento debe tener fondo, también debe cumplir unos requisitos de forma o estructura básica que lo hagan legible, comprensible y claro para todo quien lo evalúe o lo lea. Estos aspectos de forma han sido adoptados —con ligeras variaciones— por la gran mayoría de los centros académicos en el mundo, en aras de dar un orden lógico a los procesos de investigación.

Puede decirse que la comunidad académica ha establecido que toda propuesta debe contener como elementos mínimos aceptables: 1) el planteamiento general del problema, 2) la justificación de la investigación, 3) el estado del conocimiento actual sobre el tema de interés, 4) los objetivos, tanto generales como específicos, que se pretende alcanzar, 5) el plan de análisis de los datos obtenidos, 6) el cronograma que se va a seguir y 7) el presupuesto detallado del proyecto.

En términos generales, se recomienda que la extensión del protocolo, incluyendo anexos, gráficos, tablas, referencias bibliográficas y agradecimientos, sea de entre 20 y 50 páginas tamaño carta, letra Arial de 11 puntos e interlineado de 1,5, sin exceder de 62-65 caracteres por línea.

ESTRUCTURA BÁSICA DEL DOCUMENTO

Presentación del protocolo

Título del proyecto

Responde a las preguntas: ¿Qué se va a investigar y quiénes lo van a hacer?

Debe ir en la primera página del documento; debe ser corto y preciso, y en él debe quedar reflejada la pregunta de investigación; además, debe dejarle claro al lector o al revisor los objetivos y las variables principales del estudio, de suerte que estas sean las "palabras clave" para la clasificación de este. Se recomienda que el título no pase de 150 caracteres, y, como recomendación clave, debe estar concebido de forma que capture el interés inmediato del lector.

Debajo del título se debe identificar claramente a las personas que participan en el proyecto, junto con sus títulos académicos; estos últimos, escritos en siglas (**MD, MSC, PHD, etc.),** así como su filiación académica y el nombre de las instituciones que hacen parte de la investigación.

Resumen

Es un resumen estructurado que, en no más de 250 palabras, debe contener la información sobre el objetivo de la investigación, el tipo de diseño empleado, el lugar donde se llevará a cabo, la población sujeto de estudio, las intervenciones realizadas, el plan de análisis de la información, el tiempo previsto para la duración del estudio, sus costos y sus aspectos éticos.

Texto

Constituye el "cuerpo" del documento, y contiene la descripción detallada y paso a paso de todo el proyecto; debe contener los siguientes elementos:

a. **Introducción**
   Identificación del problema sobre el cual versará el estudio.

b. **Marco teórico**
   Responde a la pregunta: ¿Qué se sabe del tema?
   En este apartado se consignan el conocimiento actualizado o el estado del arte sobre el problema de interés. Es el fruto de una revisión sistemática, crítica y profunda sobre la literatura más relevante y pertinente acerca del tema de interés. Debe incluir las referencias esenciales, así como los hitos o los momentos clave en el desarrollo del conocimiento sobre la temática en cuestión; también, las controversias y los antecedentes que haya en torno a dicho tema.
   También es el espacio para destacar los vacíos conceptuales, y así perfilar el valor o el aporte que el trabajo hará.

c. Planteamiento del problema

Responde a la pregunta: ¿por qué se debe hacer el estudio?

Aquí se hace la formulación del problema, y de esta se derivan la acciones necesarias para determinar lo que se va a investigar. Constituye la justificación científica del estudio, y en él se fundamenta la necesidad de llevar a cabo la investigación propuesta.

La redacción de este aparte debe ser cuidadosa, para que sea clara y convincente. La Organización Panamericana de la Salud (OPS) recomienda la siguiente secuencia lógica para su elaboración:

- Magnitud del problema, frecuencia y distribución.
- Causas probables, y qué se conoce y se desconoce en torno a este.
- Existencia (o no) de posibles soluciones.
- Vacíos de conocimiento existentes.

Junto con el marco teórico, esta sección brinda al lector una panorámica clara y actual del problema, y refleja la solidez, la trayectoria y la documentación que tiene el grupo investigador.

d. Justificación

Responde a la pregunta: ¿para qué hacerlo?

Aquí se presentan los argumentos sobre la relevancia del estudio, el impacto y el aporte al conocimiento que se espera dar; también, los beneficios que se van a derivar en diferentes áreas (salud, informática, tecnología, bienestar social, etc.).

Debe ser planteada con argumentos convincentes, enfatizando la utilidad y la aplicabilidad del conocimiento derivado en el contexto específico o el ámbito particular donde se ejecute el proyecto; además, debe indicar la estrategia mediante la cual se va a desplegar la información obtenida a los potenciales usuarios del conocimiento.

e. Objetivos e hipótesis

Complementan la respuesta a la pregunta: ¿por qué hacer el estudio?

Los objetivos expresan los propósitos o las intenciones de la investigación. Pueden ser varios, y es conveniente jerarquizarlos. Una recomendación sana es que sean lo más aterrizados posible, pues los objetivos muy ambiciosos suelen convertirse en "piedras en el zapato", que, más bien, dificultan o hacen inviable el proyecto.

Deben ser enunciados de forma clara y explícita, pues son el norte al que apuntará todo el proyecto y definirán la forma como este se hará. Los objetivos mal planteados llevarán a resultados malinterpretados e incomprensibles.

Las buenas investigaciones suelen ser el fruto de un objetivo general bien planteado y bien definido, que traza una ruta clara e inequívoca para todo el equipo y para el análisis de los resultados.

Las investigaciones sin objetivos claros son lo que coloquialmente se llaman "expediciones de pesca", algo así como "veamos qué encontramos". Ante la situación hipotética de que el lector sea llamado a evaluar un proyecto de ese estilo, debe pensarse si se daría o no el visto bueno para su ejecución, y si se le autorizaría la destinación de recursos. Si las respuestas son negativas (no se daría visto bueno ni se le autorizarían recursos), ha quedado clara la importancia de este apartado en el protocolo de investigación.

Si por el tipo de diseño planteado hay lugar para el enunciado de hipótesis, estas deben ser también claras y precisas. Las hipótesis pueden presentarse en forma de una pregunta de investigación.

f. Métodos

Da respuesta a la pregunta: ¿cómo se hará?

Indica el camino, paso a paso, que se pretende seguir para alcanzar los objetivos planteados, los procedimientos metodológicos que se van a seguir, el orden de las actividades, las variables que se van a estudiar y el tipo de diseño más eficiente para alcanzar los resultados esperados, así como las técnicas y los procedimientos que se van a emplear en aras de lograr el objetivo, y los instrumentos o elementos para medir y recolectar la información.

En este apartado también se describen: la población sujeto de estudio; los medios que se usarán para el contacto con esta; los criterios de inclusión y de exclusión de los sujetos de estudio; el cálculo del tamaño de muestra del cual depende el poder del estudio, y las fórmulas y los supuestos asumidos para dicho cálculo.

Además, deben mencionarse las estrategias que se planea seguir ante una eventual pérdida de pacientes o de seguimiento.
Si el estudio, además, contempla una intervención, esta debe ser minuciosamente descrita.
Un punto esencial es determinar el plan de análisis de los datos y la forma como ellos serán presentados, para lo cual es muy conveniente el concurso de bioestadísticos.

g. **Aspectos éticos**
Los aspectos éticos que deben ser considerados son, principalmente:
- Que el proyecto haya sido planeado para cumplir las recomendaciones internacionales (Declaración de Helsinki, etc.) y las normas nacionales.
- Que el proyecto pasará por un proceso de evaluación y aprobación en un Comité Institucional de Investigación y Ética
- Que se garantice la protección a los sujetos, a los recursos naturales y económicos y al medio ambiente; que se garantice que si hay riesgos conocidos, estos son claramente inferiores al beneficio que se obtendrá, para las personas participantes en el proyecto.
- Que en el caso de investigación con animales, se respetarán todas las recomendaciones internacionales y nacionales para su bienestar, y que en todo momento se defenderán los derechos fundamentales de los animales participantes.
- Que se llevará a cabo un proceso cuidadoso, llamado ***Consentimiento informado***, en el cual se les explicará a los participantes todo lo relevante en relación con el proyecto, qué se hará, por qué, cuáles son los detalles de lo que se hará (medicamentos, vacunas, procedimientos), cuáles son los riesgos o las incomodidades, cuáles son los planes para mitigar o solucionar los problemas de salud que puedan surgir, cuáles son las responsabilidades de los investigadores, cuáles son los derechos de los participantes, y, entre otros aspectos, que el participante tiene absoluta libertad para preguntar cuanto quiera, que debe recibir respuestas satisfactorias, que puede decidir libremente si participa o no, y que cualquier decisión, incluido el retiro al cual tiene derecho en cualquier momento, no influirá en su atención médica ni tendrá consecuencias.
A menudo, el proceso culmina con la firma de un formato escrito, llamado también ***Consentimiento informado***, que tiene normas precisas al respecto; generalmente, estas son aplicables a escala nacional.

h. **Presupuesto**
Responde a la pregunta: ¿cuánto va a costar el proyecto?
En este acápite se deben justificar los diferentes costos, tanto desde el punto de vista salarial como por la consecución de equipos de diversa índole, los materiales o los insumos requeridos para la intervención planeada, los reactivos de laboratorio que se van a emplear, los seguros necesarios y las contrapartidas presupuestales estipuladas por el ente financiador, así como los costos de administración del proyecto.

i. **Cronograma**
Responde a la pregunta: ¿cuánto tiempo durará el proyecto?
Suele presentarse de forma gráfica, lo cual hace visibles las actividades que se van a desarrollar y los espacios de tiempo probables para su realización.
El cronograma traza la agenda de trabajo, y no puede desligarse del aspecto presupuestal. Debe buscarse el adecuado balance entre lo ideal y la factible; los cronogramas cortos podrían contar solo con bajos presupuestos, y los cronogramas muy largos podrían ser vistos con escepticismo por los patrocinadores.

j. **Documentos anexos**
Es muy probable que se deba anexar una carta de responsabilidad y aprobación del comité institucional de ética, así como cartas de compromiso por parte de los investigadores; si ellos son de la planta profesoral de una institución académica, deberán, además, contar con los permisos y los tiempos protegidos por parte de las instancias superiores de gobierno.
Dentro de los anexos se deben incorporar también las tablas o los formatos que se emplearán en el proceso de recolección de la información, los modelos de consentimiento

informado, los formularios de encuestas (si es del caso), etc.

PROBLEMAS COMUNES EN LOS PROTOCOLOS DE INVESTIGACIÓN

- Pensar más con el deseo que con cabeza fría, lo cual puede reflejarse en cronogramas optimistas, que no toman en cuenta los trámites institucionales, la importación y la nacionalización de equipos ni otras eventualidades que pueden afectar el curso esperado del proyecto.
- Incluir, por razones políticas o de conveniencia, a personas que no estarán comprometidas por completo con el desarrollo de la investigación.
- Tratar de contestar múltiples preguntas de investigación en un solo proyecto. Es probable que si esto ocurre no se llenen las expectativas del grupo ni de sus financiadores, o que en su presentación prime una visión de escepticismo sobre este.
- No hacer una revisión sistemática y profunda de la literatura ni del estado del arte del tema de interés, y de esta forma, hacer trabajos redundantes o faltos de novedad.
- Hacer un protocolo denso y poco práctico, pleno de detalles nimios, que de entrada generará, con alta probabilidad, rechazo en el evaluador.
- Como la otra cara de la moneda del punto anterior, que el protocolo sea muy ligero y carente de detalles metodológicos, lo cual también lleva al rechazo de la propuesta. Se estima que esta es la causal de rechazo más común. Debe recordarse que los recursos económicos son limitados, y que se deben dirigir a propuestas claras, viables y bien presentadas.
- Los proyectos con muy pocas referencias bibliográficas sugieren que el tema no se ha revisado a fondo, y los proyectos muy llenos de referencias sugieren, por el contrario, que no se hizo un filtro de ellas ni una lectura crítica de la información disponible. No hay un número "mágico" para hacerlo, pero se estima que alrededor 50 referencias dan solidez al proyecto y es un número razonable a la hora de hacer la verificación de todas o algunas de ellas.
- Incluir en el presupuesto rubros que no correspondan a la realidad, viajes innecesarios, equipos que van a tener una destinación diferente de la propuesta y compras que no estén plenamente justificadas generan la lógica y natural desconfianza de los entes patrocinadores.

RECOMENDACIONES GENERALES

Lo enunciado hasta acá es una simple guía para elaborar un protocolo de investigación que pretenda dar respuestas a las preguntas básicas que se plantean los evaluadores; sin embargo debe ser claro que no son normas inflexibles o una "camisa de fuerza", y que cada institución puede contar con modelos igualmente válidos. El punto crítico es que el modelo sea claro, concreto y completo.

También es pertinente decir que el protocolo puede variar en su estructura según el tipo de pregunta y según el diseño de investigación adoptado.

Algunas recomendaciones que pueden ser útiles son las siguientes:

1. **En protocolos donde se plantean experimentos** clínicos con intervenciones en pacientes:
   - Debe estar descrito el método de asignación aleatoria de los pacientes a los grupos o los subgrupos que se van a intervenir.
   - Debe estar claramente establecida y definida la medida de desenlace.
   - Deben incluirse la fórmula o los supuestos usados para calcular el tamaño de la muestra.
   - Deben estar consignadas las estrategias de análisis estadístico.
   - Deben estar contemplados los pasos que se van a seguir en caso de pérdidas de pacientes y de su seguimiento.

2. **En protocolos donde se van a evaluar características** operativas de métodos diagnósticos:
   - Debe estar claramente definido el patrón de oro contra el cual se va a evaluar el método en estudio.
   - Se debe garantizar la valoración a ciegas de los resultados de los grupos.
   - Si se propone un método que reemplaza el patrón de oro actual, deben quedar claramente establecidas las ventajas

de diferente orden que dicho método tiene, en comparación con la técnica vigente.

3. En protocolos de estudios de cohortes:
   - Debe quedar claro en qué momento se ensamblan las cohortes y que los participantes, si bien comparten el riesgo que se va a medir, aún no hayan desarrollado el evento que se va a evaluar.
   - Debe estar claro cómo será, quien lo hará y por cuánto tiempo se hará el seguimiento de las cohortes.
   - En protocolos de estudios de casos y controles:
   - Debe quedar clara la definición del "caso", así como la del "control".
   - Deben detectarse los posibles sesgos que se van a hallar en el estudio, y definirse cómo se controlarán estos o cómo se minimizará su impacto en los resultados finales.

Así mismo, debe tenerse en cuenta que cada diseño de investigación reviste sus propias particularidades, con ventajas y desventajas, y su elección depende de múltiples factores de diversos órdenes. Lo importante es optar por el que brinde mejores posibilidades de responder la pregunta de investigación; con base en ello, debe ajustarse el protocolo de manera que no queden vacíos en él, y tenga así altas posibilidades de ser llevado a la práctica para el beneficio de todos.

Lecturas recomendadas

Al-riyami A. How to prepare a research proposal. Oman Med J. 2008;23:66-9.

Merrill KC. Developing an effective quantitative research proposal. J Infus Nurs. 2011;34:181-6.

Singh S, Suganthi P, Ahmed J, et al. Formulation of health research protocol–a step by step description. NTI Bull. 2005;41:5-10.

World Health Organization (WHO). Health research methodology: A Guide for Training in Research Methods. 2nd ed. Manila: WHO; 2001.

World Health Organization (WHO). Recommended format for a research protocol [internet]. 2014 [citado 2014 dic. 10]. Disponible en: http://www.who.int/rpc/research_ethics/format_rp/en/

CAPÍTULO

13

Introducción a los diseños de investigación: selección del diseño apropiado

ÁLVARO J. RUIZ MORALES • CARLOS GÓMEZ-RESTREPO

Un diseño de investigación tiene indicaciones y contraindicaciones; ofrece ventajas en algunas situaciones, pero presenta dificultades que pueden ser insalvables en otras. Para el clínico, es fundamental conocer los principios básicos que deben guiar la selección de un diseño de investigación. Así como hay algunos antihipertensivos más apropiados en ciertas condiciones, y así como para un caso concreto de infección debe preferirse un antibiótico específico, para cada tipo de pregunta de investigación hay un diseño que puede ser el mejor.

INTRODUCCIÓN

El médico clínico que se propone responder una pregunta de investigación necesita conocer los aspectos más importantes de los diseños de investigación existentes, para escoger el más adecuado a la pregunta. Es evidente que no siempre se podrá conducir un experimento clínico; también, que un estudio sólido y bien conducido puede generar información inválida si el diseño no es apropiado para la pregunta. El diseño de investigación elegido debe ser el que ofrezca la mejor relación entre ventajas y desventajas y se ajuste mejor al ambiente donde se realizará el proyecto[1].

EJEMPLO

Un servicio de cirugía nota, con alarma, que ha aumentado el número de infecciones de la herida quirúrgica en pacientes sometidos a varicectomía, y quiere conocer las razones para ello. La pregunta, en el fondo, es sobre factores de riesgo, y uno de los diseños básicos está, precisamente, orientado a contestar esas preguntas. El cirujano, seguramente, planeará un ***estudio de casos y controles***.

Para el tratamiento del ébola, que ha producido una epidemia de difícil control en África, se desarrollaron unos anticuerpos, y mediante ingeniería genética se logró que una planta, similar a la de tabaco, produjera los anticuerpos luego de que se les introdujeron genes diseñados a ejemplares de dichas plantas, para que los usaran como propios. La combinación de anticuerpos debe evaluarse, y los científicos quieren saber si es o no efectiva y qué tan segura es. Es una pregunta de evaluación de una intervención, y uno de los diseños básicos sirve exactamente para ese propósito, y es el diseño ideal en esos casos y siempre que se quiera evaluar una intervención. Los infectólogos y los epidemiólogos están pensando, con certeza, en un ***experimento clínico aleatorizado.***

De la misma manera, el profesional de la salud que lee los resultados de un estudio debe conocer las particularidades del diseño que se empleó, lo

cual le permitirá ponderar los resultados, entender las limitaciones que pueda tener y conocer las potenciales fuentes de error que el diseño pueda haber producido, o controlado, según el caso.

EJEMPLO

Se publica un estudio en cuidado intensivo, sobre la efectividad y la seguridad del uso de trombolíticos en el tratamiento del infarto cardíaco. El autor revisó las historias clínicas de los pacientes que recibieron trombólisis durante los últimos tres años. Este diseño (observar y describir) no es apropiado para conocer efectividad ni seguridad, y, de entrada, el lector debe consultar con mucha precaución los resultados, pues no hay información sobre cómo fueron seleccionados los pacientes (puede haber errores que provengan de una selección, muy cuidadosa, de los pacientes con mayor probabilidad de beneficio o mejor pronóstico) y no se pueden comparar los resultados con los de un grupo similar que no haya recibido trombolíticos.

RELACIÓN ENTRE LA PREGUNTA DE INVESTIGACIÓN Y EL DISEÑO

Una vez establecida la pregunta de investigación, debe elegirse el diseño que mejor se le ajuste. Así como el éxito del tratamiento de una infección estriba en la mejor selección del antibiótico, la validez y la aplicabilidad de los resultados de un estudio dependen, en gran medida, de que el diseño de investigación sea correcto y de que se hayan tenido en cuenta las precauciones que cada diseño exige.

Con el diseño ya escogido, el paso siguiente es el cuidadoso desarrollo del protocolo, que, forzosamente, debe haber sido elaborado[2] ***antes*** de la recolección de datos, de acuerdo con las especificaciones y las necesidades del diseño seleccionado.

En los capítulos siguientes se revisarán en detalle los diferentes diseños de investigación, que aquí se presentan brevemente y de forma condensada, para ofrecer un panorama amplio y general del tema. Debe enfatizarse que la selección del diseño apropiado es vital para la validez de los resultados. Algunos diseños muy útiles para algunas preguntas pueden resultar inútiles (o, incluso, peligrosos) para otras, por cuanto pueden generar información con apariencia válida, pero sin valor real.

EJEMPLO

En los primeros pacientes con infarto agudo de miocardio en quienes se utilizó estreptoquinasa para hacer trombólisis se observó que había mayor mortalidad entre quienes habían recibido el costoso medicamento que entre las personas a quienes no se les administró. El estudio podría sugerir que la estreptoquinasa tiende a aumentar la mortalidad.

No obstante lo anterior, dado el carácter retrospectivo del estudio (se evaluó la administración o no del medicamento después de saberse si las personas habían muerto o no), lo que realmente se estaba registrando era el mayor uso de la costosa intervención en quienes tenían mayor gravedad de su enfermedad (y, por ende, mayor riesgo de mortalidad).

El uso de un diseño inapropiado para la evaluación de un medicamento pudo haber llevado a concluir, erróneamente, que la estreptoquinasa aumenta la mortalidad si se la utiliza en caso de infarto agudo.

CLASES DE PREGUNTAS QUE PUEDEN SURGIR



Cada pregunta de investigación necesita, según sus características principales, el diseño que se le adapte mejor y permita obtener mejores resultados.

En la atención en salud hay múltiples tipos de preguntas potenciales: bien sea sobre riesgo, pronóstico, evaluación de una intervención, utilidad de pruebas diagnósticas, aspectos económicos, árboles de decisiones o recomendaciones para la práctica clínica[3].

EJEMPLO

Al preguntarse si fumar aumenta el riesgo de cáncer de pulmón, se está haciendo la pregunta de si el hábito de fumar es un *factor de riesgo* para la neoplasia.

Un investigador quiere saber si el hábito de mascar chicle aumenta el riesgo de caries en personas con salud oral perfecta. En este caso quiere saber si hay causalidad en una exposición.

Continúa

Continuación

Hay interés en conocer la capacidad para predecir infartos cardíacos de la proteína C reactiva de alta sensibilidad. Se quiere saber, entonces, cuál es el comportamiento de una prueba diagnóstica.
La decisión de hacer tamizado para dislipidemia a todos los pacientes mayores de 20 años permitirá identificar casos no previsibles. Pero existe la duda de si es un gasto que guarda proporción con el posible beneficio, pues se les hará el examen a muchas personas, pero solo unas pocas tendrán resultados anormales. Debería hacerse un estudio de análisis económico (costo-efectividad, costo por cada caso diagnosticado).

Para cada una de esas y muchas otras modalidades de pregunta, hay un diseño que se ajusta más apropiadamente que otros. En algunos casos puede contestarse una pregunta con más de un diseño, pero, en general, cada diseño tiene capacidad específica para contestar bien algunos tipos de preguntas[4].

MODALIDADES DE INVESTIGACIÓN SEGÚN EL SUJETO DE ESTUDIO

La evaluación de impacto en la salud pulmonar de la exposición crónica al humo de leña se hace en sujetos; la utilidad de un antiarrítmico nuevo, también.

En general, la investigación en salud se hace teniendo como sujeto del estudio a personas, bien con enfermedad, o bien, sanas. A tal investigación, cuyo objeto de estudio son individuos, se la llama ***investigación primaria.***

Si un estudio concluye que el ketokonazol permite prevenir el desarrollo del síndrome de dificultad respiratoria aguda, pero aparecen dos nuevos estudios que no encuentran ese efecto, y un último estudio habla de un efecto de prevención, pero significativamente menor que el del primer estudio, hay información conflictiva, y cabe la posibilidad de que haya bajo poder en algunos de los estudios.

En este caso se puede recurrir a un diseño de investigación que no utiliza a individuos como sujetos de estudio, sino que se hace con estudios ya realizados. A ello se le conoce como *investigación secundaria*, y forma parte de los ***estudios integrativos***. Como su nombre lo explica, se trata de una familia de diseños que reúne información de investigaciones previas y la reanaliza de manera conjunta.

Estudios de investigación primaria

El experimento clínico es el diseño más popular (al menos, en la mente de la población), así como el más fácil de comprender. Adicionalmente, por los mecanismos de control que contiene, permite introducir muchos de los aspectos más importantes de la investigación.

Estudios experimentales



Un medicamento nuevo para la diabetes, agonista de GLP-1, debe ser evaluado para conocer su efectividad y su seguridad.
Se quiere saber si implementar un programa de educación sobre fibromialgia ayuda a los pacientes a mejorar sus síntomas.
Se propone un nuevo abordaje quirúrgico para un reemplazo articular.
Una sociedad de nutrición clínica quiere saber si una dieta específica sirve para lograr reducciones efectivas, seguras y duraderas de peso.

Todos estos casos tienen en común que se trata de intervenciones. El investigador va a introducir un medicamento, una terapia o una técnica quirúrgica, que no se habrían hecho si no hubiera sido por el estudio. Y todas las intervenciones en salud, sean o no farmacológicas, necesitan evaluación rigurosa. Para evaluarlas (medicamentos, terapias, técnicas quirúrgicas, etc.), existe un diseño específico, el más intuitivo a simple vista: el ***experimento clínico***.



Puede decirse que cuando se quiere evaluar una intervención (farmacológica o no) debería hacerse un experimento clínico. Y cuando se lea un experimento clínico, casi siempre se hace para evaluar una intervención.

Si se quiere evaluar un medicamento antihipertensivo, se lo utiliza en un grupo previamente definido de pacientes, en quienes se miden uno o varios desenlaces relacionados con el medicamento, y se comparan los resultados hallados con los resultados de un grupo al cual

se asigna a recibir otro tratamiento; habitualmente, el mejor del que se disponga hasta el momento. Este diseño se llama ***experimento clínico***, porque permite al investigador controlar la mayor cantidad de aspectos del estudio, como en los experimentos de laboratorio, en física, química o biología.

Quien inicia el estudio decide cuál será la población que estudiará, cuál será la intervención, cuáles serán las variables desenlace, cómo van a ser medidas, por quién y en qué condiciones. Decide también cuál será el método de asignación de los sujetos a los diversos grupos de intervención. Toma, además, decisiones sobre el método de análisis y sobre el control de otras variables importantes[4,5].

El experimento clínico como diseño tiene un enorme poder para proporcionar información válida, ya que tiene múltiples mecanismos automáticos de control de sesgos y de error. Pero el experimento debe cumplir algunas condiciones principales para poder producir resultados de calidad:

- Debe tener siempre un grupo control.
- La asignación de los sujetos a los grupos de tratamiento debe hacerse al azar.
- La evaluación de los sujetos y de los resultados debe hacerse con la menor opción de sesgos, por lo cual se recurre a enmascarar a los sujetos, a los evaluadores, y a veces, incluso, a quienes hacen el análisis.
- Finalmente, a menudo debe hacerse un análisis por grupos según la gravedad de la enfermedad, o según la clasificación por alguna variable considerada de importancia en los resultados.

Estas características que debe tener el diseño del estudio hacen que, en últimas, tenga los "apellidos" corrientes: ***experimento clínico aleatorizado***, ***doble enmascarado*** y algunos más que se agregan en la descripción, como *multicéntrico*, ***estratificado,*** etc.

La **tabla 13.1** muestra los estudios de investigación primaria que pueden utilizarse en la clínica, con un orden que sugiere la solidez de sus resultados. Como puede verse, en general, por tener el mayor control sobre los sesgos y los factores de confusión, se da al experimento clínico una posición de máxima solidez.

Entre los diseños de investigación primaria (esto es, con individuos como sujetos de investigación), el experimento clínico ofrece, entonces, las mejores seguridades en cuanto a garantía de calidad. Es el diseño más prestigioso, y, sin duda, el que más se parece a la investigación básica de laboratorio.

Necesidad de grupo control

En los experimentos clínicos debe haber un grupo control por varias razones: la primera de ellas porque, en medicina, el desenlace no siempre es predecible[4].

EJEMPLO

No discutimos la necesidad de cirugía en un paciente con apendicitis perforada, porque sabemos bien cuál es el resultado en caso de no hacerlo. No necesitamos controles para evaluar un medicamento que logre la sobrevida en un paciente con rabia comprobada, pues, hasta el momento, el resultado es invariablemente fatal.

Pese a ello, la situación contraria es más común en medicina: se desconoce el desenlace y, aunque exista una tendencia, no puede predecirse la evolución del paciente.

Tabla 13.1 Jerarquía de diseños de investigación en estudios de investigación primaria

Estudios de pruebas diagnósticas o de concordancia
• Estudios de correlación
• Estudios de concordancia
• Estudios de características operativas
Estudios de aspectos clínicos o epidemiológicos
• Estudios observacionales
– Diseños descriptivos
▪ **Reporte de casos**
▪ **Serie de casos**
– Diseños analíticos
▪ **Estudios de casos y controles**
▪ **Estudios de cohortes**
– Estudios experimentales
▪ **Experimentos clínicos aleatorizados**

EJEMPLO

En un paciente con un infarto agudo de miocardio, edema pulmonar y choque cardiogénico, la mortalidad es alta. Pero si sobrevive, no podemos saber cuál habría sido su evolución en caso de no habérsele administrado el tratamiento experimental. En este caso, solo un grupo control puede darnos esa información específica.

Selección de la intervención de control

¿Cuál debe ser la intervención al grupo control? No es lógico comparar un antibiótico nuevo contra un placebo en una neumonía por *Streptococcus* ***pneumoniae***, **pues no solo no se** obtiene información útil, por no tener con qué compararla, sino porque no hay ninguna justificación ética para administrar placebo a un paciente con neumonía bacteriana. Se debe comparar el antibiótico nuevo con otro antibiótico. Pero si el comparador no es la mejor opción (razonablemente disponible), de nuevo se obtiene información poco útil, por no tener comparación con el tratamiento habitual, y faltamos a la ética al no administrar el mejor tratamiento. La pauta para elegir la intervención o el tratamiento control es administrar el mejor tratamiento que haya hasta el momento, en condiciones reales[6].

Uso del placebo

Se utiliza placebo en los casos donde no está probada la efectividad de una intervención, ***y*** *solo en esos casos*. **Un placebo es una sustancia** que puede, eventualmente, producir cambios, pero no por un mecanismo específico y conocido de acción.

EJEMPLO

En un niño con epilepsia, aun si hay medicamentos muy prometedores en desarrollo, es imposible, desde el punto de vista ético, comparar uno de los nuevos candidatos contra un placebo. Sin embargo, en un paciente con vasculitis no se está obligado a dar bloqueadores de canales de calcio a todas las personas con padecimientos similares, pues no hay certeza de su eficacia, de manera que un grupo puede recibir placebo.

La recomendación ética es que se reserve el uso de placebo para casos donde no hay un tratamiento con demostración de efectividad y seguridad. Aunque hay excepciones:

EJEMPLO

En pacientes con epilepsia, en tratamiento con dosis apropiadas de un medicamento en quienes se logran reducciones sustanciales del número de crisis, pero aún hay persistencia de estas, y se quiere evaluar la adición de otro medicamento antiepiléptico.
En el control de dolor neuropático por diabetes, un medicamento de base no es suficiente, y se quiere evaluar el uso de una combinación con otra medicación.
En los dos casos anteriores, los pacientes reciben un medicamento de base, y se quiere evaluar la posibilidad de adicionar una nueva opción. En este caso, a un grupo se le agrega el nuevo medicamento y al otro grupo se le agrega placebo, para mantener el enmascaramiento. En estos casos se justifica el uso de placebo, para evaluar la adición, pero ya los pacientes tienen un tratamiento de base.

Asignación de los sujetos de investigación

La decisión de quién va a cada uno de los brazos de un experimento (medicamento nuevo o placebo, por ejemplo) puede terminar siendo desigual e inclinada a un extremo (sesgada) si se asignan a la intervención nueva los pacientes de más alto riesgo. Esa desigualdad significa que el grupo de la intervención tiene peor pronóstico, y desde antes del inicio del estudio se puede prever que les vaya peor en el estudio.

Debe, entonces, evitarse el riesgo de asignar a los pacientes artificialmente y, quizá, de forma sesgada.



Dejar la asignación en manos del azar, pero con técnicas y procedimientos explícitos y correctos, permite abolir la posibilidad de sesgos en la asignación y ofrece muchas otras ventajas, pues si hay diferencias entre los grupos después de la asignación, estas no se deberán a sesgos, sino al azar[7].

Y es altamente posible (en particular, en tamaños de muestra que no sean pequeños) que las condiciones conocidas y no conocidas de los sujetos queden repartidas de manera simétrica y comparable (por ejemplo, la edad, el sexo, la prevalencia del tabaquismo, antecedentes de bajo peso al nacer, historia de depresión, etc.).

Enmascaramiento o cegamiento

Un investigador convencido de la eficacia o de la seguridad de un medicamento puede influir en los resultados al evaluar la respuesta clínica, aunque la asignación haya sido hecha al azar. El paciente también puede quedar impresionado por la maniobra, y sentir mejoría por su percepción del tratamiento. Para evitar que su percepción influya en los resultados, se puede enmascarar al paciente (***estudio enmascarado o ciego***). También se acostumbra enmascarar al evaluador (***estudio doble enmascarado***), lo cual evita los sesgos en la evaluación. Y se habla también de enmascarar a quien hace el análisis de los datos (***estudio triple enmascarado***); de nuevo, para lograr los resultados más puros posibles, sin influencias externas.

Estratificación

En estudios entre pacientes con diversos estadios de enfermedad, el nivel de gravedad puede influir en los resultados.

Un paciente con hipertensión estadio 1 puede responder mejor a un inhibidor de enzima convertidora de angiotensina 2 (IECA), por ejemplo, que otro paciente con hipertensión estadio 2. La separación de los pacientes según su nivel de hipertensión desde el comienzo del estudio para hacer un análisis por grupos, según el estadio de hipertensión, ayuda a controlar la influencia de la gravedad de la hipertensión en los resultados del estudio. A eso se le llama *estratificación*, y es una forma de controlar los aspectos que pueden resultar importantes en el desenlace.

Un experimento clínico que reúna los elementos mencionados (es decir, que sea controlado, cuya asignación haya sido aleatorizada, que sea doble o triple enmascarado y que tenga análisis estratificado —esto último, si es apropiado, según el caso—) es la mejor garantía de que, en lo posible, se están evitando problemas, pues se aprovechan las ventajas del diseño[2].

Limitaciones para la realización de experimentos clínicos

La mayoría de las limitaciones existentes se relacionan con dificultades logísticas para llevar a cabo un experimento, o con limitaciones éticas.

Para evaluar si aumenta la accidentalidad cuando se conduce y se habla por celular simultáneamente (sin sistema de manos libres), podría planearse un experimento clínico que aleatoriza a un grupo destinado a utilizar el sistema manos libres, y a un grupo control, a no utilizarlo. Sin embargo, hay múltiples limitaciones logísticas: no será fácil garantizar que haya suficiente exposición, que ésta sea simétrica entre los grupos, que haya exposición coincidente con los accidentes que se presenten, etc.

Se quiere saber si la exposición a la radiación aumenta el riesgo de leucemia. No hay ninguna justificación ética para asignar a un grupo a vivir en un área contaminada por un accidente nuclear, por cuanto se estaría exponiendo a los sujetos a otros riesgos, ya conocidos y confirmados, de la radiación. La limitación ética impide realizar dicho estudio.

En casos como los anteriores, o en casos similares, no es fácil, por razones logísticas, hacer un estudio (pues sería muy difícil lograr que un grupo se someta a una intervención como mascar chicle si no lo hace habitualmente, o no hacerlo si está acostumbrado, o hablar por celular y no utilizar manos libres, etc.), o por razones éticas (ya que no sería correcto exponer a un grupo a radiación para saber si aumenta el riesgo de leucemias, ni a un medicamento de cuyos efectos tóxicos se tiene una alta sospecha). En estos casos, debe recurrirse a un diseño diferente. Lo que se hará es observar lo que ha ocurrido, o lo que ocurrirá, sin que el investigador intervenga y proponga él una intervención. A esos diseños se los llama, por tal motivo, ***estudios observacionales***.

Estudios observacionales

EJEMPLO

La unidad de cirugía laparoscópica de un hospital universitario quiere dar a conocer la experiencia que se ha tenido en los dos años de su existencia. Se hace un informe con el número de pacientes operados, los tipos de cirugía, los promedios de edad, las complicaciones que se presentaron, el promedio de duración de las intervenciones, etc.

Un otorrinolaringólogo quiere saber si es más común que haya pérdida de audición en los jóvenes que escuchan habitualmente música con audífonos que en quienes oyen música sin estos. Se toman, entonces a quienes oyen música con audífonos (pacientes

Continúa

Continuación

expuestos) y se los compara con quienes no usan audífonos (no expuestos) y se evalúa su audición luego de un tiempo de exposición.
Para saber cuáles son las posibles causas de miopía, se hace un estudio en el que se toma a personas con miopía (casos) y se los compara con personas sin miopía (controles) y se les pregunta por los factores de los cuales se sospecha que pueden estar relacionados: antecedente familiar de miopía y el "trabajo de cerca", como ocurre con la lectura habitual.

No hay una intervención para el estudio. Se describe lo que ocurre (como en el caso de la cirugía laparoscópica), y en ocasiones no solo se lo describe, sino que se pueden comparar grupos; por ejemplo, los fumadores con quienes no lo son, o los pacientes que tienen cáncer con quienes no lo tienen. En el último caso citado, por ejemplo, se evalúa cuántas veces más común es que sean muy lectores los pacientes miopes cuando se comparan con los no miopes.

Estos diseños se clasifican en dos grupos: 1) estudios ***descriptivos*** (en ellos solo se describe, como en el caso de la cirugía laparoscópica) y 2) estudios ***analíticos*** (en ellos se comparan dos grupos y se hace un análisis de las diferencias, como en los casos de miopía). En los primeros sobresalen el reporte de un caso y la serie de casos; en los segundos, los estudios de casos y controles y los estudios de cohortes.

Diseños descriptivos

Reporte de un caso

El reporte de caso es el tipo más sencillo de diseño en la investigación clínica. Se limita a describir cuidadosamente un caso observado[8], y por ello se lo clasifica, dentro de los estudios observacionales, como un estudio descriptivo. Buenos ejemplos de reporte de caso son: la información sobre un efecto colateral raro de un medicamento; la asociación poco común entre dos enfermedades; la descripción de una enfermedad rara, o la de manifestaciones raras de una condición.

Relativamente a menudo, la observación es el resultado de una mente inquieta, profunda y fresca, que puede llevar al descubrimiento de un mecanismo de enfermedad o de una asociación, o que permite perfeccionar el conocimiento sobre un mecanismo. La descripción en detalle de un caso puede llevar al conocimiento profundo de algún aspecto básico; por lo general, un mecanismo farmacodinámico o un aspecto fisiopatológico específico. Esto hace del reporte de caso una fuente útil de información sobre aspectos de las ciencias básicas que puede ser útil en las ciencias clínicas.

También puede suceder que el caso sea interesante, pero no proporcione información de gran utilidad, por tratarse de un fenómeno con baja probabilidad de repetición: por ejemplo, en el caso de la coincidencia en un mismo paciente de tres tumores diferentes. La descripción no deja de ser atractiva, pero las conclusiones prácticas no son muchas, pues la probabilidad de volverse a encontrar con un caso similar es de casi cero. En tales circunstancias, incluso, puede llevar al lector desprevenido a sacar conclusiones inexactas.

Ante un caso raro, el problema no radica en saber si puede ocurrir, sino en saber si aparece más a menudo que lo que cabe esperar tan solo por efecto del azar[9]. Esta pregunta no puede responderse con un reporte de caso. La **tabla 13.2** muestra las características de los reportes de casos.

Serie de casos

La serie de casos hace la descripción de eventos observados en un grupo de pacientes. La tabla 1 muestra la ubicación de las series de casos en los estudios observacionales (en contraposición con los experimentales) y descriptivos. La experiencia de un urólogo con sus últimas 35 litotripsias, o los 65 pacientes tratados a lo largo del último año por pancreatitis hemorrágica en el hospital X son ejemplos de series de casos.

Las series tienen las mismas limitaciones que el reporte de casos: se conoce la frecuencia de aparición del desenlace de interés en el grupo descrito, pero no es posible saber si este es o no representativo de la población, de manera que no pueden calcularse frecuencias para la población. Se tiene, además, el numerador de la

Tabla 13.2 Características del reporte de caso

• Descripción detallada de datos fisiológicos, clínicos y paraclínicos
• Útil para la detección de casos raros
• Puente entre investigación clínica e investigación básica
• Fuente de hipótesis
• Estímulo para hipótesis posteriores

ecuación para el cálculo de la incidencia, pero no el denominador, por lo cual la información no es representativa de la fuente de origen. La **tabla 13.3** resume las características de las series de casos.

Los diseños observacionales descriptivos no permiten establecer relaciones causales, pues no permiten saber si la posible causa apareció antes que el efecto y no tienen grupo control[2,9].

En la recolección de los casos hay grandes probabilidades de sesgos, bien porque se trata de los pacientes altamente seleccionados de un hospital terciario, o bien, por ser muy raros y curiosos, lo cual los hace más interesantes y atractivos, pero también les disminuye la probabilidad de que de allí se obtenga conocimiento generalizable[4]. Estos pacientes de casos raros tienen más probabilidades de ser hospitalizados, de que sus casos sean comentados en reuniones clínicas y, definitivamente, muchas más de que se publique acerca de ellos[10].

Los estudios de esa clase son inmejorables fuentes de hipótesis, y, por tratarse de descripciones completas, pueden llevar a una mejor comprensión de los mecanismos de las enfermedades o de su comportamiento. La actitud del clínico ante los estudios observacionales descriptivos debe ser crítica y cuidadosa, aunque también de interés, por cuanto pueden ser fuente de conocimientos al producir preguntas, generar hipótesis, o ser acicate para la realización de estudios. La **tabla 13.4** resume las características de los estudios descriptivos.

Otros estudios

Se han mencionado algunos de los elementos que dan solidez a un estudio. Solo la presencia de un grupo control permite evaluar si el resultado observado en el grupo de estudio depende del factor en evaluación[11]. Obtener mejoría en el 87 % de los casos en el grupo de intervención nada dice en tanto no se sepa el porcentaje de mejoría en el grupo control, al que no se le sometió a la intervención. La selección de los pacientes debe estar exenta de sesgos y los grupos deben ser comparables, y mediante pruebas estadísticas se mide la probabilidad de que los resultados hayan sido producidos por el azar.

Ocasionalmente, se hace necesario estudiar una posible causa en una enfermedad en la que se reconoce una multicausalidad, y ello obliga al investigador a exigirle al diseño que pueda controlar las posibles causas diferentes de la de interés. La **tabla 13.5** resume estas necesidades, que pueden satisfacerse con diseños más elaborados. Dentro de los estudios observacionales, la inclusión de un grupo control evita muchos de los problemas ya mencionados y permite plantear y estudiar hipótesis. Dichos estudios se llaman ***analíticos***, por su capacidad para analizar la información obtenida.

Diseños analíticos

Estudios de casos y controles

EJEMPLO

A raíz de la publicación de varias series de casos sobre malformaciones fetales (específicamente, ausencia de extremidades), se hizo un estudio en el que se tomó a un grupo de mujeres cuyos hijos habían nacido durante el último año (1961) con malformaciones congénitas, y se compararon sus datos con los de un grupo de mujeres con características similares a las anteriores, pero cuyos hijos, considerados normales, habían nacido en la misma época y en los mismos hospitales. Había un grupo de casos (mujeres con hijos anormales) y un grupo de controles (mujeres con hijos normales).

Tabla 13.3 Características de las series de casos

• Descripción atemporal
• Grupo reducido, y a menudo, altamente seleccionado
• No proporcionan información sobre prevalencia ni sobre causalidad
• Generadoras de hipótesis

Tabla 13.4 Características de los estudios descriptivos

Ventajas
• Útiles en la descripción de casos raros
• Generadores de hipótesis
• Estímulo para la realización de estudios
Desventajas
• Alto potencial de sesgos
• Ausencia de grupos de comparación
• Imposibilidad para la medición del azar

Tabla 13.5 Elementos necesarios en el estudio clínico

• Grupo control
• Comparabilidad en los grupos
• Control de factores diferentes del estudiado
• Medición de efecto del azar (pruebas de significación)

La **figura 13.1** muestra las características del estudio, la esencia de los casos y controles; primero se reúne a los pacientes, y luego se investiga retrospectivamente la exposición al factor de riesgo. La **figura 13.2** muestra las características del estudio.

Por la dirección en el tiempo para recolectar información respecto a la exposición, se los denomina ***estudios retrospectivos***. No es raro, sin embargo, que con esta expresión se conozcan también las revisiones de historias, que son, realmente, series de casos. La descripción es correcta, pero se refiere a la dirección de la recolección de los datos, ya que parte del desenlace y busca hacia atrás en el tiempo las exposiciones; se refiere, entonces, a la dirección, y no es el nombre del diseño. El nombre de ***estudio de casos y controles*** es el nombre de este diseño, y debe evitarse el uso de la palabra ***retrospectivo*** como sustituto del nombre.

Este es el diseño analítico más popular. En enfermedades poco comunes o con tiempos de latencia prolongados (por ejemplo, el cáncer de pulmón y tabaquismo) son particularmente útiles, pues facilitan encontrar a pacientes que hayan desarrollado la enfermedad, por medio de registros hospitalarios o de registros nacionales. Es recomendable que los pacientes entren al estudio a medida que se van presentando: ello permite tener un modelo estandarizado y probado para recolectar la información, capaz de asegurar que la información requerida va a estar presente[1-3].

EJEMPLO

Para estudiar la relación entre fumar cigarrillo y el cáncer de pulmón, se toma a 450 pacientes con cáncer de pulmón (casos) y a 450 sujetos de características muy similares a los anteriores, pero sin cáncer de pulmón (controles). Se busca la información sobre los factores de interés de los cuales se quiere saber si tienen o no alguna asociación con la enfermedad estudiada. Se buscan: la frecuencia de antecedente de cáncer en la familia, el historial de tabaquismo y los antecedentes de exposición a tóxicos o a otros contaminantes ambientales.

Una vez obtenida la información, se utiliza la tabla de 2 × 2, o de cuatro casillas, que se muestra en la **figura 13.3**. De esta casilla, es útil la información de la frecuencia de la exposición en cada uno de los grupos según la enfermedad, así como la frecuencia de la enfermedad en cada uno de los grupos según la exposición. No es útil, sin embargo, la frecuencia de la enfermedad, que se ha producido artificialmente al hacer la recolección de los pacientes (se tomaron 450 casos y 450 controles; luego, la frecuencia de la enfermedad sería, artificialmente, del 50 %).

El cálculo del riesgo relativo indirecto (*odds ratio*, en la literatura anglosajona) permite conocer cuántas veces más riesgo de haber estado expuestos tienen quienes padecen la enfermedad.



Figura 13.1 Estudio sobre malformaciones fetales y talidomida.

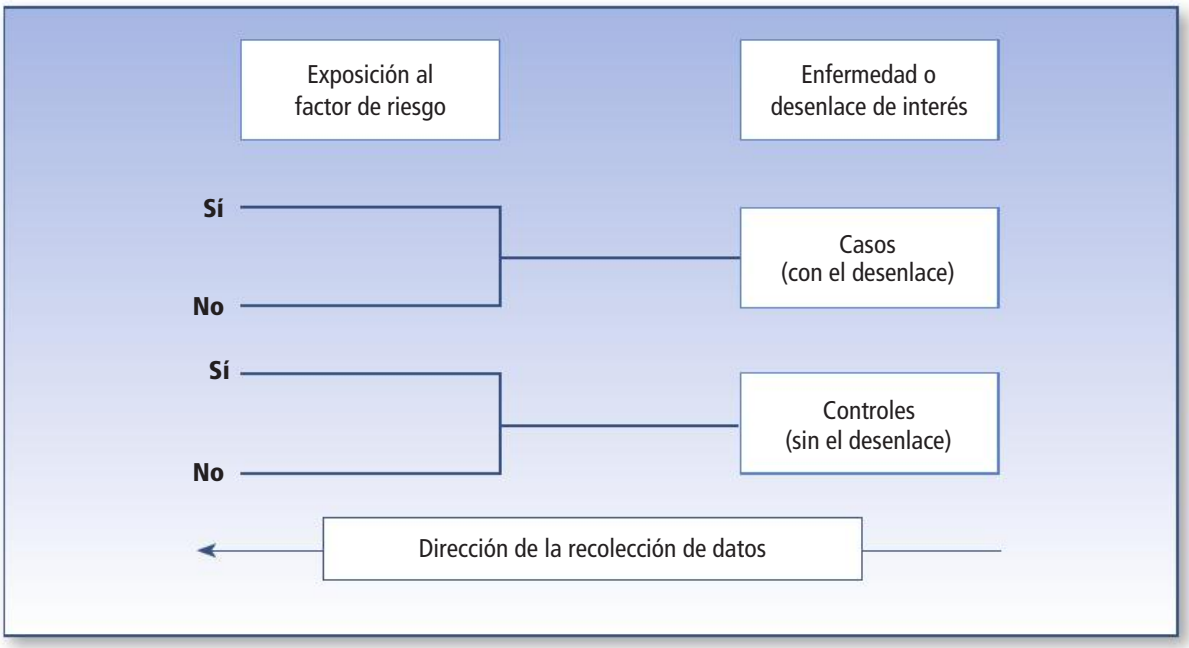


Figura 13.2 Estudio de casos y controles.

En el caso del cigarrillo y el cáncer del pulmón, el riesgo relativo indirecto informado en la literatura es de 16. Quiere decir que en quienes tienen cáncer de pulmón es 16 veces más probable haber estado expuesto al cigarrillo que en quienes no tienen cáncer de pulmón.

Las muchas ventajas del estudio de casos y controles lo hacen uno de los diseños más populares: se ha mencionado que por partir de personas que ya desarrollaron la enfermedad de interés se evitan las dificultades para la recolección de los pacientes, y que esto lo hace un buen diseño para enfermedades raras[1].

Además de lo anterior, el mecanismo ágil para obtener a los pacientes y la búsqueda retrospectiva de la información de estudios así los hacen poco costosos, y puede afirmarse que, en términos generales y en comparación con otros diseños sólidos, son los estudios más económicos.

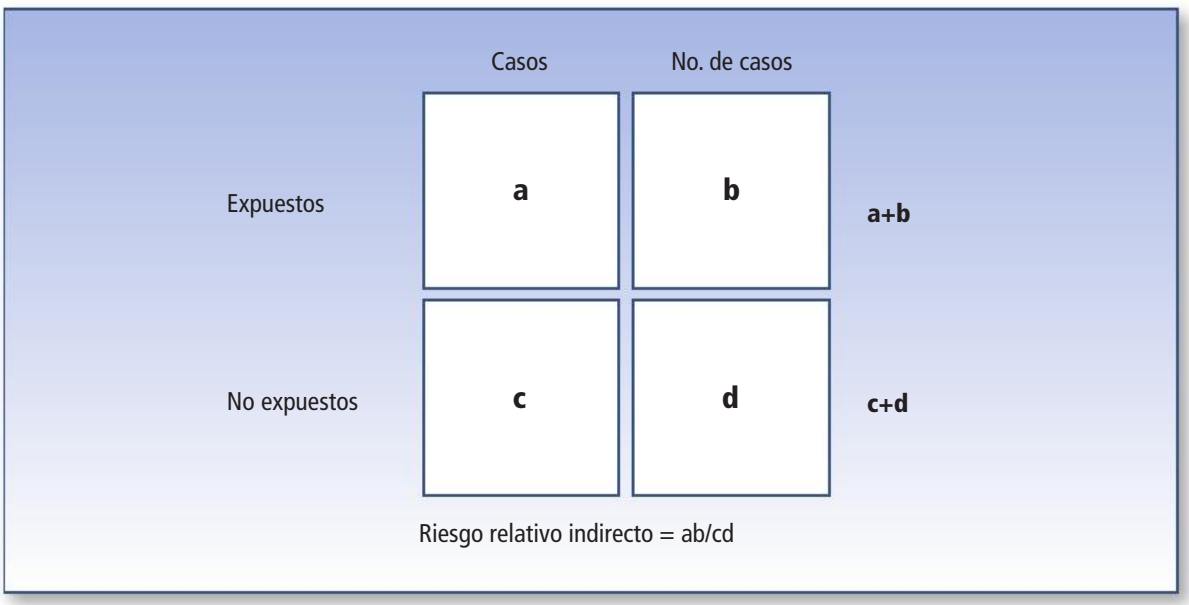


Figura 13.3 Evaluación de la exposición en estudio de casos y controles.

La **tabla 13.6** enumera, así mismo, otra ventaja: al empezar el estudio con el desenlace de interés, se evita el tiempo de latencia, que a menudo es de varias semanas o meses, como en el caso de la nefritis intersticial por antiinflamatorios no esteroideos (AINE), o, incluso, de varios años, como en los casos de la mayoría de agentes carcinógenos o el desarrollo de enfermedad coronaria.

En cuanto a las desventajas, ya se ha comentado acerca de la susceptibilidad a los sesgos de este tipo de estudios[1,2]. Las razones principales son:

1. Los grupos no ocurren naturalmente, lo que le permite al investigador manipular los grupos de comparación.
2. La exposición se mide después de que ha ocurrido la enfermedad.

Para evitar los problemas relacionados con la agrupación artificial, los pacientes deben haber tenido la misma oportunidad de hallarse expuestos.

Este diseño proporciona buena información sobre la asociación entre el desenlace que se ha escogido y uno o varios factores de riesgo.

EJEMPLO

Se buscan factores de riesgo para cáncer de seno (desenlace) en un estudio de casos (mujeres con cáncer de seno) y de controles (mujeres sin cáncer de seno). Se pueden investigar múltiples factores (edad, historia familiar de cáncer de seno, varias mutaciones genéticas específicas (BRCA1, BRCA2, ATM, TP53, etc.), raza, enfermedad fibroquística, radiación torácica previa y uso de estrógenos.

Pese a lo anterior, aunque el diseño puede informar sobre múltiples factores de riesgo, no da información sobre otros desenlaces. Pueden investigarse múltiples factores de riesgo para malformaciones fetales (como irradiación o medicamentos tomados durante el embarazo, antecedentes familiares de defectos similares, factores ambientales, etc.), pero solo para el desenlace de interés: malformaciones fetales.

Por otra parte, en la mayoría de los casos, el diseño no permitirá conocer la relación temporal entre el factor de riesgo y el desenlace[2,12].

No podrá saberse si primero fue la ingesta de la droga antiemética, y luego, a consecuencia de lo anterior, las malformaciones; o si, por el contrario, la presencia de las malformaciones condicionó de alguna manera el exceso de náuseas y el vómito en las pacientes, lo que los llevó a ingerir más a menudo la talidomida.

Un diseño cuidadoso y un método estandarizado, crítico y probado para recolectar la información pueden evitar la mayoría de los problemas comunes en los estudios de casos y controles. No hay un análisis estadístico, por sólido o ingenioso que sea, capaz de corregir la presencia de sesgos graves en un estudio[11].

Estudios de cohortes

EJEMPLO

Se estudia a un grupo de mujeres que trabajan en salones de belleza, y que tienen exposición constante a tinturas para el pelo, y se las compara con mujeres de características similares, pero que no tengan la exposición. Se las sigue en el tiempo, para ver si hay diferencias en cuanto al desarrollo de cáncer de vejiga.

Si quiere hacerse énfasis en la causalidad, hay un diseño que permite obtener dicha información, y permite también evitar algunos de los problemas comunes con los estudios de casos y controles. Se toma a un grupo de sujetos expuestos a un potencial factor causal (como en el ejemplo, a tinturas) y a otro grupo no expuesto, y se busca en el tiempo quiénes desarrollan el desenlace. A este diseño se le llama ***estudio***

Tabla 13.6 Características de los estudios de casos y controles

Ventajas
• Sirven para estudiar enfermedades raras
• Son relativamente económicos
• Obvian las latencias largas en estudios de causalidad y de pronóstico
Desventajas
• Generalmente, no permiten establecer causalidad
• Son muy susceptibles a sesgos: – De selección – De memoria – De búsqueda

de cohortes. **Para que la única diferencia sea** inicialmente la de la exposición, se necesita que los grupos sean tan parecidos entre sí como sea posible. A esto se le llama un ***estudio de cohortes*, tomando prestado un término de la** milicia romana, donde la cohorte (***cohors*) era** una unidad táctica del ejército, entrenada para actuar "como un solo hombre". La **figura 13.4** muestra el diseño básico de ese tipo de estudio. Cabe anotar que la recolección de los datos se hace hacia delante en el tiempo, de forma prospectiva, pues parte de la exposición hacia el desenlace[4,13].

Este tipo de estudio tiene algunas características que lo hacen excepcionalmente útil en investigación clínica: sirve para determinar de forma directa la incidencia (es decir, el *riesgo absoluto*), por cuanto se parte de personas expuestas y se busca el desarrollo de la enfermedad. La exposición, además, se explora sin la posibilidad de introducir sesgos, pues se averigua antes de que aparezca la enfermedad[11].

El mismo estudio puede buscar la relación entre la exposición escogida y muchas enfermedades. Además, en su planteamiento sigue la lógica clínica: dado que la persona está expuesta, ¿desarrollará la enfermedad?

Algunos factores limitan su utilización y lo hacen un poco menos popular de cuanto merecería por sus ventajas: el seguimiento de los pacientes toma mucho tiempo en algunos casos (el tiempo de latencia de la enfermedad puede ser prolongado, como ocurre con los carcinógenos, o con los factores de riesgo para enfermedad coronaria). Además, en términos generales, requiere inversiones económicas mayores que en el resto de diseños, no solo por los costos de seguimiento (actualización de información, visitas y contacto permanente con los sujetos), sino por los costos de las evaluaciones (visitas, entrevistas y exámenes médicos paraclínicos). Además, los estudios de esa clase son relativamente ineficientes, pues necesitan a un número elevado de sujetos para lograr una frecuencia aceptable en el evento de interés, y solo pueden evaluar la relación entre la enfermedad y la exposición a pocos factores; de hecho, los registrados ***al comienzo* del estudio.** La **tabla 13.7** resume los aspectos positivos y negativos a ese respecto[2,12,14].

También hay posibilidades de sesgos en este diseño; es decir, puede haber desviaciones sistemáticas al recolectarse la información.

A pesar de sus dificultades logísticas, el tiempo y el costo, este diseño es una de las mejores armas del clínico, por su enorme valor en el estudio de causalidad y porque permite estudiar los efectos de exposiciones de interés.

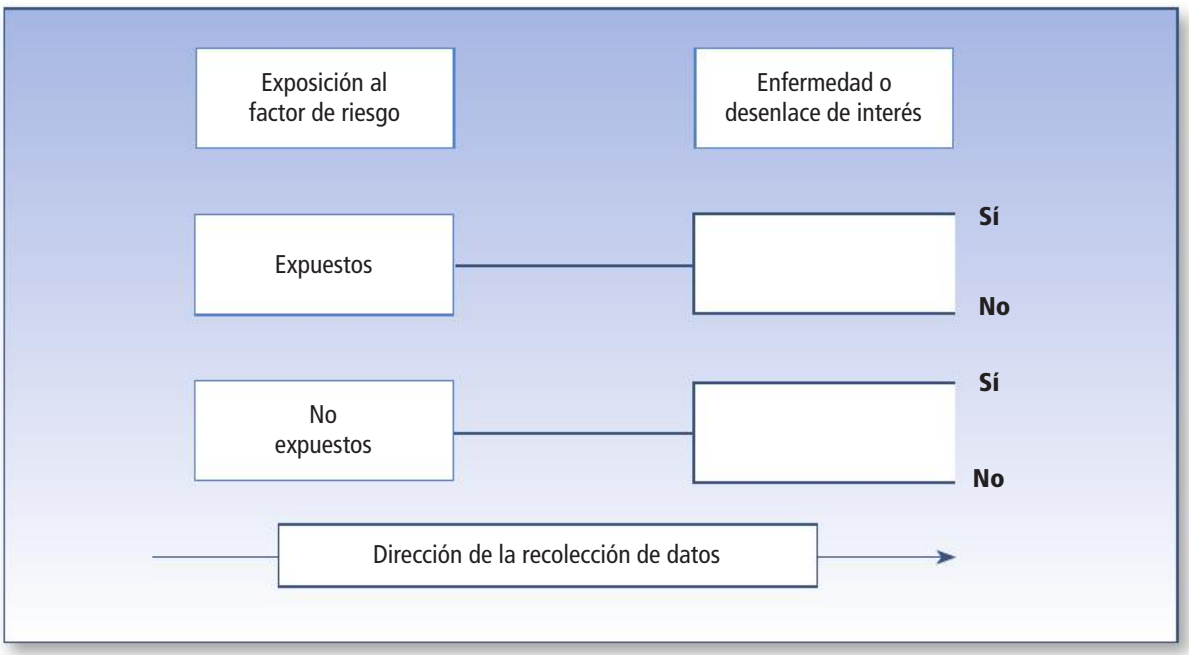


FIGURA 13.4 Estudio de cohortes.

Tabla 13.7 Características de los estudios de cohortes

Ventajas
• Permiten cuantificar incidencias
• No hay sesgos a la hora de identificar la exposición
• Pueden evaluar efectos de la exposición en múltiples desenlaces
• Siguen la lógica clínica (de exposición a la enfermedad)
Desventajas
• Pueden tener un alto costo económico
• Su tiempo de estudio es prolongado
• Pueden requerir un alto número de pacientes
• Pueden evaluar simultáneamente pocos factores

Estudios de concordancia

Estudios de correlación

EJEMPLO

Se sabe que el aumento de grasa intraabdominal se asocia al aumento en el riesgo de desarrollar diabetes y al aumento en el riesgo de enfermedades cardiovasculares (a esto se lo conoce como *riesgo cardiometabólico*). Para medir la grasa intraabdominal, se debe recurrir a un escanograma, lo que significa dificultades logísticas y económicas para la implementación de esta medida en la evaluación de riesgo. Lo ideal sería tener una medición alterna, fácil, que se relacione estrechamente con la grasa intraabdominal. Se establecieron por escanografía los valores de grasa intraabdominal en un grupo de pacientes, y se les midió el perímetro abdominal con una cinta métrica: una medición fácil y rápida. Además, se evaluó si había relación entre las dos variables. Se encontró que los valores más bajos de una medición se correlacionaban con los valores más bajos de la otra, y que los cambios en una variable se correlacionaban muy bien con cambios en la otra. La evaluación permitió proponer la medición de manera rutinaria de perímetro abdominal, pues sus resultados se correlacionan muy bien con los de adiposidad intraabdominal.

Cuando se quiere entender la relación entre dos variables o predecir cambios en una variable a través de cambios en otra (saber si aumenta o no la presión arterial cuando aumenta el peso), se recurre a estudios de correlación[4,15]. Como en el ejemplo, puede saberse si el aumento en una variable se correlaciona con el aumento en otra, o si la relación es inversa. Se sabe, por ejemplo, que a medida que aumentan los valores de triglicéridos disminuyen los valores de colesterol de alta densidad (HDL, por las siglas en inglés de *high density lipoprotein*). Esta es una correlación negativa, pues el aumento de triglicéridos y el aumento de peso tienen una correlación positiva. Tal estrategia no responde, sin embargo, a la pregunta de si un método diagnóstico puede reemplazar a otro, para lo cual se debe recurrir a estudios de concordancia.

Estudios de características operativas

Hay otros tipos de diseño especialmente relacionados con exámenes diagnósticos. A menudo, el médico clínico quiere saber cuál es la utilidad de un nuevo método diagnóstico, (colecistoscopia para la exploración de la vía biliar, biopsia de tiroides para diagnóstico de una masa, ecografía de vías biliares para cálculos, etc.). Dicho valor se mide en ***sensibilidad***; es decir, la capacidad para identificar lo anormal como anormal, y como ***especificidad***, que es la capacidad para identificar lo normal como normal.

Estos conceptos se conocen como *características operativas de un examen*[12,15], y su estudio debe afrontarse de una manera diferente de cómo se afrontan los estudios anteriores: debe decidirse, en primer lugar, cuál será la "verdad" contra la cual se comparará la nueva técnica.

Se habla de ***certeza diagnóstica*** cuando se tienen piezas histológicas o un método diagnóstico incontrovertido, al que, a su vez, se le llama ***estándar de oro***.

Una vez decidido el estándar de oro, debe diseñarse el estudio para que tanto los pacientes con la entidad que se va a estudiar como las personas sin ella sean incluidos[16]. De la información obtenida de las ecografías en los pacientes con la entidad, por ejemplo, con cálculos en la vesícula (según la cirugía; en este caso, el estándar de oro), se puede determinar la sensibilidad: del número total de pacientes con cálculos (a + c, en la **figura 13.5**) se puede determinar el porcentaje de los enfermos que tuvieron un resultado anormal (positivos).

Para el cálculo de la especificidad, necesariamente debe contarse con la información de

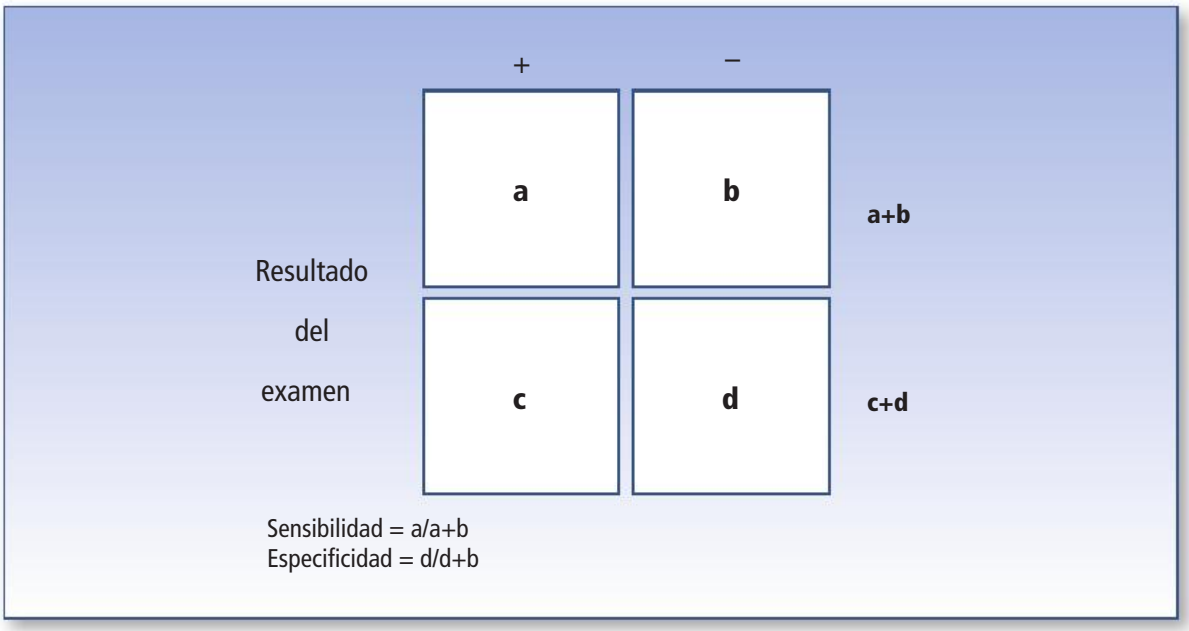


FIGURA 13.5 Características operativas.

cómo se comportó el examen en las personas sin cálculos. El porcentaje de (b + d) que haya tenido un resultado negativo es la especificidad.

Para estudios así se necesitan cuatro elementos:

1. **Un estándar de oro claramente definido, y** que, idealmente, incluirá tejidos o será el mejor método disponible para diagnóstico.
2. **Pacientes que, según el estándar de oro,** tengan la enfermedad buscada, y personas que no la tengan.
3. **Los resultados del examen investigado, tanto** en los pacientes con la enfermedad (según el estándar de oro) como en quienes no la tienen.
4. **Un número de pacientes, determinado de** antemano (tal como sucede para los otros diseños). La regla que aún rige respecto a calcular el tamaño de la muestra para esta clase de estudios es: "la calidad de los resultados depende de que el número final de pacientes sea adecuado".

Estudios de tendencia, de correlación y de concordancia

Al comparar los resultados de dos o más métodos para hacer mediciones, es habitual que sus resultados no sean dicotómicos (es decir, positivo o negativo, presente o ausente), sino datos continuos: 7,5 mmol/dl de glucosa, 4,3 mmol/dl de colesterol, etc. Una primera aproximación es la de describir la relación entre dos variables asociadas, lo cual puede hacerse mediante un índice de tendencia.

EJEMPLO

Se quiere describir la variación de los triglicéridos (en mmol/dl) según el peso, en categorías de 5 kg. El índice describiría la tendencia de los triglicéridos, los cuales aumentan a medida que aumenta el peso. Las pruebas disponibles para este tipo de evaluación informan sobre la relación mutua entre las variables (si la hay), su dirección (*positiva* si es directa, y *negativa* si el aumento de una de las variables se acompaña de disminución de la otra) y su magnitud.

Si se comparan las técnicas A y B para medir el colesterol, la prueba de correlación dirá si, en términos generales, los valores altos en la prueba A se acompañan de valores altos en la prueba B, y viceversa. Sin embargo, puede suceder que el método A, inicialmente diseñado para mediciones en suero, se esté utilizando en sangre total, y sus resultados sean consistentemente mayores que los reales por efecto de pigmentos presentes en la sangre.

Los datos de A serán siempre mayores que los de B, pero la diferencia será siempre constante, pues la interferencia en las mediciones hecha

por A estará siempre presente y en la misma intensidad. Como resultado, tendremos que A y B tienen una excelente correlación: tanto los valores bajos como los medios y los altos de A se acompañan de valores correspondientemente bajos, medios y altos de B. Sin embargo, los valores de A estarán siempre un poco por encima de los de B.

Igual podría suceder con un tensiómetro descalibrado, que marca siempre 10 mmHg por encima del valor real. La comparación entre dicho tensiómetro y uno bien calibrado mostrará una buena correlación entre los valores, ya que también aquí los valores bajos, medios y altos lo son con cualquiera de los dos tensiómetros. Sin embargo, los valores serán siempre mayores en el caso de nuestro tensiómetro defectuoso.

Según lo anterior, se hace necesaria una prueba que muestre no solo si la tendencia es la misma, sino si los valores comparados producen los mismos resultados. A eso se le llama ***prueba de concordancia***.

Un estudio de concordancia[15] mide la extensión en la que los resultados producidos por dos técnicas diferentes son iguales entre sí.

Según el estándar de oro que se utilice, pueden distinguirse dos tipos de estudios de concordancia: 1) los ***estudios de conformidad***, donde hay un estándar de oro claramente identificado, como en el caso de la prueba de un método rápido para la determinación de glucosa por tiras reactivas, que se compara con el método bioquímico estándar; y 2) los ***estudios de consistencia***, **cuando no se asume ninguna** de las dos técnicas como la ''correcta''; generalmente, porque no existe un estándar de oro reconocido.

Estudios integrativos

También llamados ***estudios secundarios*** cuando se hace alusión a las revisiones sistemáticas de la literatura y al metaanálisis, reúnen información proveniente de estudios previos. Se trata de diseños de investigación, formales y con requerimientos de calidad y cumplimiento, como los otros diseños[17].

Revisiones sistemáticas de la literatura y metaanálisis

Desde hace mucho tiempo se ha acostumbrado hacer revisiones de temas, para lo cual a menudo se hace una revisión de la literatura existente sobre el tema. En estos casos, toda la información hallada es tenida en cuenta, con al menos dos posibles fuentes de error: 1) si no se hace una búsqueda con método, puede no encontrarse información de gran importancia. Y 2) si no se tienen criterios de evaluación de los hallazgos, puede ocurrir que se asigne la misma importancia a la información proveniente de investigación de óptima calidad y a una opinión no fundamentada.

Por lo anterior, aparecieron los ***diseños formales de investigación***, **que responden a estas** inquietudes: una revisión completa y exhaustiva, con reglas explícitas de búsqueda y que busca responder a una pregunta concreta, corresponde a una revisión sistemática de la literatura, en contraposición a las revisiones narrativas[17].

En algunos casos se hace la revisión formal, así como una apreciación crítica de la información, y con criterios explícitos se reúnen los datos provenientes de estudios homogéneos y de óptima calidad (por lo general, experimentos clínicos aleatorizados), para hacer un análisis final de toda la información reunida. A este diseño formal se le conoce como ***metaanálisis***.

EJEMPLO

El tratamiento de la hipertensión arterial en los pacientes mayores de 65 años suscitó gran controversia en el pasado; en particular, cuando se trata de hipertensión sistólica aislada. Dado que se dudaba de su utilidad y de que hubiera un balance adecuado entre riesgos y beneficios, se hicieron múltiples experimentos clínicos, cuyos resultados eran conflictivos. Uno de los metaanálisis ahora disponibles de dichos estudios demostró claramente que, al reunir los datos de los experimentos de buena calidad llevados a cabo, se justifica muy claramente tratar la hipertensión, tanto la sistodiastólica como la sistólica aislada, en las personas mayores de 65 años.

Otros modelos

En otros casos se hace necesario reunir la información proveniente de investigaciones

con fines específicos o para hacer un análisis particular. Es el caso de los estudios que buscan conocer las relaciones entre costos, riesgos y beneficios[18].

Estudios de análisis económico

Si se busca responder a la pregunta de si es adecuado el balance entre riesgos y costos, por un lado, y beneficios por otro, los estudios de análisis económico son de ayuda[19]: permiten conocer, en términos cuantitativos, cuánto es el beneficio, en términos prácticos clínicos, de una inversión (cuánto cuesta salvar, en términos monetarios, un año de vida al hacer un tratamiento farmacológico de las hipertensión arterial), cuál es el beneficio, en términos económicos, de una intervención, también medida en términos económicos (costo-beneficio del tratamiento hipolipemiante) o la comparación entre dos o más intervenciones, para saber cuál ofrece la misma cantidad de años de vida de calidad por el menor costo (estudios de costo-utilidad).

Análisis de decisiones

Integrar la información preexistente que considera no solo los desenlaces, sino los costos, de diferentes alternativas permite ofrecerle al clínico una guía para tomar decisiones ante escenarios diversos, con sus resultados en términos de salud y económicos[18,20].

Guías de práctica clínica

Una vez obtenida la información existente en la literatura y en otras fuentes de información, y evaluada cuidadosamente con criterios estandarizados, se puede ofrecer una guía de uso general, que resuma la información encontrada y le dé una calificación a cada una de las fuentes de evidencia, así como calificar la recomendación que de ellas se desprenda[21].

Tienen enorme utilidad para guiar la práctica clínica por la calificación de las recomendaciones, pero debe tenerse muy claro que una guía de práctica clínica es siempre el comienzo de una decisión, consensuada con el paciente, y no el final. La guía pretende, justamente, eso: guiar al clínico, y no puede ni debe ser utilizada para reemplazar el juicio clínico.

MÉTODOS CUALITATIVOS

EJEMPLO

La hipertensión arterial es una de las causas más importantes y prevalentes de infartos cardíacos, de infartos cerebrales, de insuficiencia cardíaca y de complicaciones renales y en la retina. Los medicamentos para tratarla son cada vez más precisos; se conocen datos a partir de investigación sobre su efecto en la reducción de las complicaciones, y sobre las indicaciones para hacer combinaciones. Hay también más y mejores medicamentos, mejores combinaciones, precios más al alcance del consumidor y sistemas de salud preocupados por su adecuado control. Y, sin embargo, no se ha logrado un control apropiado de esta enfermedad. Al buscar razones que dificulten la efectividad y la eficacia de los medicamentos del caso, surgen inquietudes de interés: ¿por qué los pacientes no toman los medicamentos? ¿Por qué no tienen adherencia a ellos, o a las medidas no farmacológicas, como bajar de peso, mantener un peso adecuado y cuidar la alimentación? Estas preguntas, cuya respuesta es tan relevante para la salud de individuos y poblaciones, tienen un tinte diferente de las respondidas hasta ahora. Y necesitan, claramente, un enfoque distinto.

A menudo, el objeto de un proyecto tiene que ver con percepciones del sujeto de investigación; con discursos o palabras como suceden en la vida real; con observaciones en el territorio del sujeto; con su lenguaje y sus textos, y con explicaciones que se dan desde la subjetividad de la persona. Más que la frecuencia de un fenómeno, el interés de un estudio puede estar en la explicación de este, o de su percepción por parte del individuo. En este caso, no solo es correcto tener un diseño apropiado, sino que sería inútil intentar alguno de los diseños anteriores.

Los métodos cualitativos son apropiados para responder preguntas como qué es esto o aquello, cómo y por qué se presenta algún desenlace o enfermedad, qué representación simbólica o significado tiene para las personas, cómo varía en determinadas circunstancias o cuál es la naturaleza del fenómeno individual y social.

Sirven también para tratar de explicar la acción social mediante métodos interpretativos.

Los métodos cualitativos contribuyen significativamente a la investigación clínica y de servicios de salud; en especial, cuando se procura profundizar sobre temas o problemas determinados que requieren enfoques especiales.

Estos métodos se han ido haciendo cada vez más naturales en la investigación en salud (particularmente, a raíz de la publicación de una significativa serie de artículos publicados en el ***British Medical Journal*** (BMJ) acerca del tema —y que, a su vez, dieron origen al libro *Qualitative research in health care*—). Claramente, no solo no son herramientas opuestas a los estudios cualitativos, lo cual significaría una pérdida para ambas tendencias[22], sino que tienen utilidad mucho más allá de sus campos iniciales y naturales[23], como la antropología, la sociología, la psicología y la psiquiatría.

Muchas preguntas en las ciencias de la salud necesitan enfoques cualitativos: por ejemplo, las mencionadas acerca de no adherencia o no aceptación de tratamientos farmacológicos o de estilos de vida. También, preguntas como las razones de las dificultades para implementar medidas con el fin de disminuir los riesgos; o por qué, a pesar de conocerse el riesgo de algunas enfermedades, su conocimiento no llevaba a cambios en el comportamiento, como en el caso del uso del cigarrillo y su relación con los cánceres de pulmón, de páncreas, de vejiga y otros; o por qué no se logra prevenir mejor las enfermedades de trasmisión sexual o los embarazos indeseados, pese al conocimiento sobre la efectividad de los preservativos y otros métodos idóneos para ello.

Otras preguntas podrían ser: indagar cuál es la historia de los pacientes suicidas, las razones para recurrir a medicinas alternativas o las explicaciones para las diferencias en cuanto a la frecuencia de remisión de pacientes entre instituciones con niveles similares de complejidad y acceso. Estas y otras situaciones sirven para ejemplificar la utilidad de las mencionadas metodologías.

Con el fin de indagar acerca de este tipo de preguntas, se han empleado diversas técnicas, como entrevistas en profundidad o a grupos focales, que permiten indagar con mayor detalle las razones o los motivos esbozados para un determinado comportamiento, así como métodos de observación (participante o no) con el fin de describir o comprender la conducta de los profesionales de la salud en diferentes ámbitos.

Ahora bien, los métodos cualitativos y sus técnicas de investigación para obtener datos de las personas y de su ambiente suelen ir ligadas a una perspectiva teórica, que permite formular modelos explicativos sobre lo indagado, así como informar e interpretar los datos de la investigación. Entre dichas perspectivas o enfoques epistemológicos se cuentan la etnografía, el interaccionismo simbólico, el constructivismo, la etnometodología y la fenomenología, que pueden ser útiles para guiar la recolección y el análisis de información, y permiten, aunque haya aún diferencias en su valoración[24], facilitar que los comprendan terceros observadores.

Un uso asiduo de las mencionadas estrategias es la comprensión inicial de ciertos fenómenos. Sirven, por ejemplo, para conocer las palabras y las categorías que servirán de base para elaborar cuestionarios o escalas, los cuales serán utilizados para conocer la opinión de personas y cuantificar características sociodemográficas, personales, conductas, riesgos, actitudes, aptitudes, creencias, preferencias u otras de interés. Sirven también para validar hallazgos del uso anterior (a lo que se llama ***triangulación de datos***) y para interpretar o tener explicaciones más profundas u holísticas de lo generado mediante métodos cuantitativos.

Se hará enseguida una presentación breve de las estrategias más comunes para estudios cualitativos, y de sus rasgos principales.

Observación

Este tipo de técnica permite ver u observar a las personas del estudio en su propio medio. En ocasiones, se hace de manera participante (observación participante), donde el investigador se introduce (***se incorpora***) en el diario vivir de las personas a quienes está observando, participa de todas sus actividades e incorpora sus propias costumbres, lo cual le permite recuperar un gran número de experiencias que obtiene al convivir con la población. Son clásicos a este respecto los ***estudios etnográficos***; por ejemplo, los realizados por antropólogos que viven durante un tiempo en comunidades y posteriormente describen e interpretan lo que encontraron en torno a las preguntas formuladas.

En otros casos se hace la ***observación no participante*** de lo que está ocurriendo: por ejemplo,

cuando un investigador (observador entrenado) va a un área del hospital y observa la conducta de los profesionales de la salud ante las exigencias de los familiares, o de los pacientes en la sala de espera, en consulta, etc.

Entrevistas

Las entrevistas son diálogos individuales o en grupo.

Entrevistas individuales

Se hacen con las personas de acuerdo con un libreto que puede ir desde muy estructurado hasta muy abierto, con preguntas que le permiten al entrevistado una mayor asociación al tema tratado. Las entrevistas pueden ser: 1) ***estructuradas*, y seguir un libreto totalmente establecido;** o 2) ***semiestructuradas*, y tener una serie de** ítems, dominios o categorías para explorar, lo cual permite profundizar con nuevas preguntas, como reacción a lo que se vaya encontrado, o 3) entrevistas ***abiertas***, en las cuales se suscita una conversación con alguna pregunta que evoque la temática planteada, y donde se sigue la línea de asociación de la persona entrevistada profundizando mediante preguntas que permitan ahondar en la temática de investigación. Ello permite abrirse a los nuevos hallazgos en la medida en que ocurren, y evitar quedarse con los planteamientos explicativos o hipotéticos que preliminarmente se tenían.

Entrevistas grupales (grupos focales)

Esta técnica se utiliza en un grupo de personas (habitualmente, entre 6 y 14), que se reúnen con un moderador para dar sus opiniones acerca de un tema. Por lo general, en esta técnica se tienen dos o tres temas (preguntas) que pretenden suscitar la opinión de todos los participantes. En este método, más que llevar a cabo una discusión, se profundiza acerca de todas las opiniones que los participantes tengan sobre un tema. Se debe ser muy cuidadoso en la selección de los participantes y en la moderación, así como en dar importancia a todas las opiniones y en moderar el grupo para obtener libertad y diversidad. Como ejemplo, mediante un grupo focal podría conocerse la opinión de profesionales de la salud sobre políticas en salud, o sobre las dificultades para implementar guías de práctica clínica en un país.

Mediante estas técnicas de entrevista (individual o grupal), se obtienen textos de las conversaciones, que se evalúan con técnicas de análisis de datos y programas como AtlasTi, InVivo u otros, guiados por teorías o categorías previamente establecidas; o, como ocurre en la mayoría de los casos, mediante categorías no previamente definidas, sino emanadas de los datos recolectados; o mediante técnicas mixtas (categorías previas, y otras emergentes de los datos).

Basados en documentos, análisis de textos o discursos

En estas técnicas se provee un análisis de textos (libros, historias, narraciones, relatos biográficos, videos, grabaciones, etc.) que pueden servir para ahondar en explicaciones o en factores asociados a determinados desenlaces en salud. Como ejemplo, en una población pequeña se presentan casos de reacciones a la vacuna para el virus del papiloma humano (VPH). Se duda de si dichas reacciones están o no asociadas causalmente a la vacuna, y para su estudio sería de mucha utilidad el análisis de las comunicaciones radiales, de televisión o de prensa al respecto. Esos relatos y su análisis podrían proveer explicaciones plausibles acerca de cómo y por qué se ha presentado una reacción repetida, aunque confinada localmente, con más visos de ser una respuesta psicológica masiva y sin que parezca haber una explicación causal. Otro uso común de este método es en el análisis de textos de psicoterapias, psicoanálisis u otras terapias.

Otras técnicas

Existen otros métodos cualitativos, como las técnicas de consenso, la técnica Delphi y los grupos nominales, que sirven para llegar a opiniones comunes entre grupos de expertos o personas que estén trabajando acerca de un tema común. Tales métodos serán abordados en este libro, para llegar a recomendaciones en la realización de guías de práctica clínica.

Otros métodos, como el estudio de caso, podrían complementar el reporte de caso comúnmente empleado en ciencias de la salud (y mayormente, en salud pública), como medio para profundizar en problemas o situaciones de salud. Finalmente, otras técnicas como la investigación-acción permiten hacer inter-

venciones en salud a largo plazo y retomando los resultados obtenidos en el transcurso de la investigación como insumos para las etapas subsiguientes.

Cabe anotar que todos los métodos mencionados requieren técnicas precisas, que, a su vez, exigen planeación, lógica y su empleo riguroso para recolectar e interpretar datos.

CONCLUSIONES

La solución a cada pregunta clínica debe ser resuelta con el diseño en particular que se le ajuste mejor. Para cada pregunta hay al menos un diseño ideal: por ejemplo, para informar sobre un paciente o sobre un grupo de pacientes, los reportes y las series de casos cumplen el propósito.

Para la búsqueda de factores de riesgo, el *estudio de casos y controles* es lo ideal, así como para la mayoría de estudios que evalúan causalidad el mejor diseño es el ***estudio de cohortes***. La evaluación de un medicamento o de una intervención deberá hacerse siempre con un *experimento clínico*. Para conocer las características de una prueba diagnóstica, la ***sensibilidad*** y la ***especificidad*** entregan esa información. Si se quiere saber si una variable puede predecir otra, los ***estudios de correlación*** son los indicados; pero si se quiere saber si un método diagnóstico puede reemplazar a otro, deberá hacerse un *estudio de concordancia*.

La existencia de resultados no consistentes en varios estudios (por lo general, en experimentos clínicos) puede resolverse con un ***metaanálisis***. Y los aspectos económicos, cada vez más importantes en la atención en salud, pueden ser estudiados con los diseños de minimización de costos, costo-efectividad, costo-beneficio y costo-utilidad. Cuando se quiera indagar acerca de cómo y por qué se presenta una situación en salud, así como la opinión de sus participantes, resulta conveniente utilizar alguno de los métodos cualitativos según la conveniencia.

El conocimiento de los diseños, de sus indicaciones y de sus fortalezas, así como de sus debilidades, le permitirá al investigador hacer la selección apropiada para su diseño, de acuerdo con el tipo de pregunta que se tenga.

Referencias

1. Rothman KJ, Greenland S, Lash TL. Case-control studies. En: Encyclopedia of quantitative risk analysis and assessment. Chichester, UK: John Wiley & Sons; 2008.
2. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: The essentials. 5th ed. Baltimore: Williams & Wilkins; 2012.
3. Haynes RB, Sackett DL, Guyatt GR, et al. How to do clinical practice research clinical epidemiology: How to do clinical practice research. Boston: LWW; 2005.
4. Miettinen OS. Epidemiological research: An introduction. Dordrecht Heidelberg: Springer; 2012.
5. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 4th ed. Nueva York: Springer; 2010.
6. Piantadosi S. Clinical trials: A methodologic perspective. 2nd ed. Nueva Jersey: John Wiley & Sons; 2005.
7. Tal J. Strategy and statistics in clinical trials: A non-statisticians guide to thinking, designing and executing. Waltham, MA: Academic Press; 2011.
8. Cohen H. How to write a patient case report. Am J Health Syst Pharm. 2006;63:1888-92.
9. Parfrey P, Barrett B. Clinical epidemiology: Practice and methods. Nueva York: Humana Press; 2009.
10. Laake P, Benestad HB, Olsen BR. Research methodology in the medical and biological sciences. Londres: Academic Press; 2007.
11. Hulley SB, Cummings SR, Browner WS, et al. Designing clinical research. 4th ed. Philadelphia: Lippincott Williams & Wilkins; John Wiley & Sons; 2013.
12. Riegelman RK. Studying a study and testing a test: Reading evidence-based health research. 6th ed. Philadelphia: John Wiley & Sons; 2012.
13. Polgar S. Introduction to research in the health sciences. 5th ed. Philadelphia: Churchill Livingstone; 2007.
14. Hebel JR, McCarter RJ. Study guide to epidemiology and biostatistics. Burlington, MA: Jones & Bartlett Publishers; 2012.
15. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther. 1981;29:111-23.
16. Miettinen OS. Epidemiological research: Terms and concepts. Dordrecht Heidelberg: Springer; 2011.

17. Tonelli M, Hackam D, Garg AX. Primer on systematic review and meta-analysis. En: Parfrey P, Barrett B. Clinical epidemiology: Practice and methods. Nueva York: Humana Press; 2009.
18. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. Nueva York: Oxford University Press; 2005.
19. Sloan FA, Hsieh CR. Health economics. Cambridge, MA: MIT Press; 2012.
20. Petitti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine. Oxford: Oxford University Press; 2000.
21. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine: How to practice and teach it. 4th ed. Elsevier; 2010.
22. Green J, Thorogood N. Qualitative methods for health research. Londres: Sage; 2013.
23. Pope C, Mays N, (Eds). Qualitative research in health care. Londres: BMJ Books; 1999.
24. Hammersley M. The dilemma of qualitative method: Herbert Blumer and The Chicago tradition. London: Routledge; 1989.

CAPÍTULO

14 Estudios de corte transversal

ANDRÉS DUARTE OSORIO

INTRODUCCIÓN

Los estudios de corte transversal son diseños de investigación donde las variables de interés se miden en una población o en una muestra poblacional, en un punto determinado en el tiempo. En la literatura, se referencian como encuestas de prevalencia, encuestas transversales, estudios de corte, estudios de prevalencia o estudios transversales.

Se utilizan muy a menudo en investigaciones epidemiológicas y en servicios y sistemas de salud, con el propósito de determinar la frecuencia, la magnitud y la distribución de eventos de salud o de enfermedad, en un momento y un lugar determinados, en una población, en subgrupos poblacionales o en servicios de salud. Tal tipo de estudios transversales corresponde a estudios descriptivos. También se utilizan para explorar factores asociados a alguna condición de salud o de enfermedad en la población y para probar hipótesis simples, como, por ejemplo, la diferencia de prevalencias entre subgrupos poblaciones o en una misma población, en diferentes periodos. En ese caso, se los reconoce como ***estudios analíticos***.

Un estudio de corte transversal conducido para calcular la prevalencia se denomina ***estudio de prevalencia*** (Rothman). Su utilidad se centra en que describe el fenómeno de salud y determina su frecuencia a escala poblacional. Regularmente, se llevan a cabo mediante la aplicación de encuestas (denominadas ***encuestas de prevalencia***) que pueden involucrar a la totalidad de la población (censos) o a muestras representativas de ella.

Los estudios de corte transversal que tienen componente analítico permiten calcular asociaciones entre diversas variables. Sin embargo, dado que todas ellas se exploran en un mismo momento en el tiempo, sin poderse identificar cuándo aparecieron o se adquirieron, no permiten establecer relaciones de temporalidad entre ellas, ni, por tanto, permiten inferir etiologías ni causalidades. En estos aspectos, la utilidad de dichos hallazgos se centra en aportar información para generar nuevas hipótesis de investigación, las cuales, a su vez, deben probarse con otros diseños.

La interpretación de las asociaciones halladas en estudios transversales debe incluir los potenciales sesgos (***desviaciones sistemáticas***) que las expliquen. Por ejemplo, si una empresa les hace una encuesta a sus trabajadores y en-

cuentra que en un tipo de labor específica los trabajadores presentan más a menudo dolor lumbar, podría asumirse que dicha labor incrementa la frecuencia de dolor lumbar. Sin embargo, para establecer que la variable catalogada como exposición (el tipo de labor del trabajador) se relaciona con la condición o el problema de salud catalogado como desenlace (el dolor lumbar), es necesario demostrar que dicha exposición precedió al desenlace, lo cual no es posible con tal diseño. En este ejemplo, puede ocurrir que la labor responda a una reasignación laboral, a consecuencia de un dolor que previamente se habría presentado, y, por ende, dicha labor no se relaciona causalmente con el dolor.

Diferentes autores mencionan que la limitación de los estudios de corte transversal para identificar las relaciones de temporalidad y contribuir en la inferencia de causalidad se puede compensar con la flexibilidad y la rapidez para explorar asociaciones entre múltiples exposiciones y múltiples efectos.

En el presente capítulo se describen los aspectos generales del diseño de estudios de corte transversal, el análisis de los resultados, sus ventajas, sus limitaciones y algunos aspectos relacionados con su reporte y su publicación.

Diseño

Los aspectos metodológicos más relevantes para considerar en el diseño de un estudio de corte transversal se resumen en los siguientes puntos:

Consideraciones éticas

Al igual que con los demás diseños de investigación, los estudios de corte transversal deben conducirse según los lineamientos éticos internacionales (Declaración de Helsinki, Council for International Organizations of Medical Sciences —CIOMS—, etc.) y las normas éticas regionales y nacionales, y según los principios éticos de beneficencia, no maleficencia, autonomía y justicia social. Dentro de dichas consideraciones, debe cuestionarse la pertinencia misma de la investigación. Muchos estudios se conducen tratando de resolver preguntas ya resueltas en la misma población, y, eventualmente, con diseños de mayor jerarquía. Es importante cuestionar la pertinencia y la utilidad potencial del estudio, así como el uso de los recursos empleados y el respeto a los recursos.

Pregunta de investigación



Para los estudios de prevalencia, la pregunta es de carácter cuantitativo-descriptivo. Si se pretende indagar por factores asociados a una condición clínica o una enfermedad, la pregunta será de carácter analítico.

Las limitaciones para establecer relaciones de temporalidad hacen que tal tipo de diseño no contribuya a resolver preguntas relacionadas con etiología, causalidad o factores de riesgo. Y cabe enfatizar que una de las principales utilidades de ese tipo de estudios es generar hipótesis.

Población blanco

Se refiere a la población finita, con características o atributos clínicos y demográficos definidos, a la cual podrán generalizarse los resultados del estudio.

Pertinencia de realizar el estudio con la población total o en una muestra

Los estudios de corte transversal pueden involucrar a la totalidad de la población, como ocurre en los censos o en los estudios de servicios de salud específicos, o bien, pueden llevarse a cabo a partir de muestras, que deben ser representativas de la población y, preferiblemente, seleccionadas de forma aleatoria.

Las razones para hacer un estudio con la totalidad de la población, o en una muestra de ella, son similares para los demás diseños de investigación, y se discutieron en el **capítulo 5** "Población y muestra".

Representatividad de la muestra



Si se quiere determinar la prevalencia de una condición partiendo del estudio realizado con una muestra, ella debe ser representativa de la población. Si se establecen criterios de selección basados en la presencia de condiciones o de posibles exposiciones, la prevalencia encontrada no será la poblacional, sino la de la población seleccionada.

Al tomar muestras representativas de la población, se favorece la posibilidad de hacer un componente analítico exploratorio de fac-

tores asociados. Sin embargo, generalmente el tamaño de muestra se establece para estimar la prevalencia, con una precisión específica y con un poder y nivel de confianza determinados. Eso limita el poder del estudio para hallar asociaciones estadísticamente significativas entre los factores asociados.

Selección de variables y escalas de medición

Deberán incluirse las variables demográficas y clínicas pertinentes a resolver la pregunta de investigación. La escala de medición empleada para cada una de ellas dependerá del propósito del estudio.

Elaboración y validación de los instrumentos y las técnicas de medición

Generalmente, se utilizan encuestas o cuestionarios que permitan identificar o registrar las variables clínicas y demográficas de interés. Las técnicas para elaborar y validar dichos instrumentos se discutirán en el **capítulo 19** "Estudios cualitativos".

Fuentes de datos

Cuando se utilizan fuentes primarias, cada individuo de la población o de la muestra seleccionada será valorado, bien sea a través de una entrevista, una encuesta, un examen físico, exámenes de laboratorio u otro tipo de valoraciones. La valoración debe conducir, en primera instancia, a determinar la presencia o no de una condición o una enfermedad prevalente, o bien, a que se pueda seleccionar a los individuos con alta probabilidad de presentarla. En este caso se harán pruebas confirmatorias. En segunda instancia, las valoraciones permitirán identificar otros factores que pueden relacionarse con la presencia o la ausencia de la enfermedad o de la condición prevalente.



A menudo, los estudios de corte transversal se adelantan con fuentes secundarias, como, por ejemplo, registros poblacionales, resultados de encuestas de base poblacional o diversos registros institucionales. Si esto sucede, deberá hacerse siempre un análisis de calidad de los datos, el cual, si es pertinente, se utilizará para ajustar procesos estadísticos y para discutir los resultados.

Análisis

La **figura 14.1** presenta el esquema general de un estudio de corte transversal que pretende identificar no solo la frecuencia de la condición o la enfermedad, sino, además, los factores asociados.

Al igual que con los demás diseños de investigación, el análisis dependerá de los objetivos planteados, de las variables seleccionadas y de la escala de medición empleada para cada una de ellas **(tabla 14.1)**.

El análisis de información suele iniciar con el cálculo de estadísticas descriptivas. Dicho análisis permitirá conocer las características generales de la población o la muestra estudiada y estimar prevalencias de las enfermedades o las condiciones, al igual que de los factores asociados. La descripción suele hacerse mediante distribución de frecuencias (número de sujetos u observaciones dentro de cada categoría de la variable), frecuencias relativas (distribución porcentual de las observaciones dentro de las categorías de la variable) y proporciones.

Si el objetivo del estudio es determinar la prevalencia, a partir de los conteos de personas enfermas se establecerá la proporción dividiendo el número de casos encontrados por el total de la población. Si el estudio pretende comparar prevalencias entre subpoblaciones, se calcularán las proporciones utilizando numeradores y denominadores acordes con los grupos de interés definidos. Pero si, además, se pretende explorar factores asociados, deberán discriminarse los datos de acuerdo con la presencia o la ausencia de enfermedad y la presencia o no de algún factor, y, a partir de estos datos, calcular las medidas de asociación.

Una tabla tetracórica permitirá organizar los resultados de la siguiente forma:

Para determinar la asociación entre la ocurrencia de una enfermedad y un potencial factor de riesgo, se recurre al cálculo de medidas de asociación. Si se tratara de establecer el exceso de riesgo de enfermar asociado a algún factor, se debería, en primer término, calcular las tasas de incidencia tanto de los pacientes expuestos como de los no expuestos a dicho factor. A partir de ellas, se debe calcular como medida de asociación la razón de tasas, también denominada ***razón de riesgos*** o ***riego relativo***. Estas medidas de incidencia solamente pueden obtenerse en estudios en los cuales se parte de

Tabla 14.1 Lista de verificación para planear un estudio de corte transversal

N.°	Tema	☑
1	Definir claramente los criterios diagnósticos de la enfermedad, el rasgo o la condición de interés.	
2	Definir la población sujeto de estudio.	
3	Definir la estrategia de muestreo; preferiblemente, probabilística.	
4	Establecer un plan de análisis.	
5	**Si es pertinente, determinar el número de encuestadores necesarios según el número de** encuestas requeridas, su duración prevista y la densidad de la población.	
6	Establecer un cronograma.	
7	Elaborar el cuestionario, incluyendo todas las variables necesarias para el análisis.	
8	Conformar equipos de trabajo de campo.	
9	Entrenar a coinvestigadores y a encuestadores.	
10	Llevar a cabo una prueba piloto del cuestionario y de los formatos de recolección de datos.	
11	Ajustar el formulario de acuerdo con los resultados de la prueba piloto.	
12	**Si se trata de una encuesta poblacional:** a. Levantar un mapa del área y numerar las viviendas. b. Llevar a cabo una visita de reconocimiento de las viviendas seleccionadas. c. Considerar la publicidad por medios masivos de comunicación. d. Determinar la necesidad y la conveniencia de contactar a líderes comunitarios. e. Establecer las funciones de los supervisores de campo, y su papel en el control de la calidad de la información. f. Establecer una logística de revisión de formatos en el campo de trabajo. g. Establecer los mecanismos para el traslado de los formatos a la central de digitación.	
13	Si se requiere, adecuar el lugar para la realización de examen físico o la toma de pruebas.	
14	Garantizar la seguridad de los formatos recolectados.	

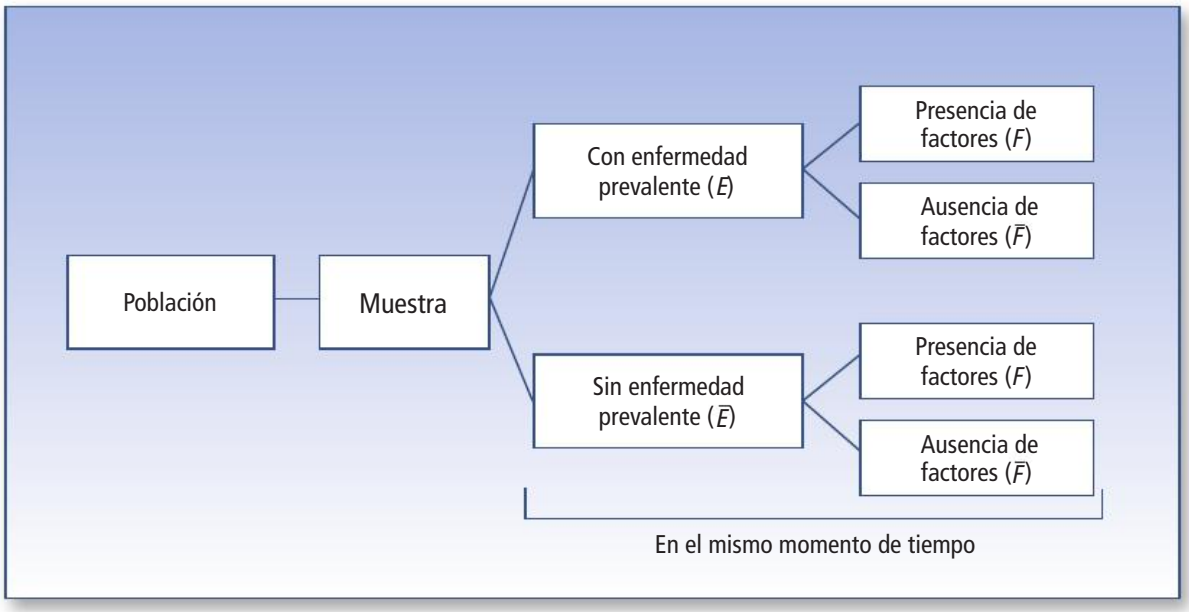


Figura 14.1 Estudios de corte transversal.

Fuente: Elaboración del autor.

individuos no enfermos y entre quienes en el tiempo se identifique y se contabilice el número de casos nuevos.

Los estudios de corte transversal, al hacerse en un momento en el tiempo, no permiten identificar los casos nuevos, y, por ende, no permiten establecer las tasas de incidencia. Sin embargo, sí permiten estimar la prevalencia.

Los casos identificados pueden tener diferentes periodos de duración, e incluyen los casos nuevos.

Para establecer la asociación entre la enfermedad o la condición de interés y cierto factor o factores, se emplea el cálculo de razones de prevalencia y razón de suertes (***odds ratio***). Estas dos medidas se calculan de forma diferente, y difieren de forma notoria en cuanto a su interpretación. Dado que no es posible establecer relaciones de temporalidad entre la presencia del factor y la aparición de la enfermedad o la condición, dichos factores no pueden considerarse factores de riesgo, y, en general, resulta inadecuado interpretarlos de ese modo. Únicamente cuando las exposiciones exploradas no se modifican en el tiempo (como, por ejemplo, en el caso del sexo, la raza, el grupo sanguíneo o la carga genética) los factores asociados que se hallen corresponderán a factores de riesgo.

La razón de prevalencias (**RP**) se calcula dividiendo la prevalencia de enfermedad en los individuos en quienes el factor se halla presente sobre la prevalencia de enfermedad en los individuos en quienes el factor no está presente. Para la **tabla 14.2**, la RP se calculará como: [(a/(a+b))/(c/(c+d)].

El valor obtenido se interpretará en términos de cuántas veces es más probable que los individuos en quienes el factor está presente muestren al mismo tiempo la enfermedad o la condición, respecto a los individuos en quienes el factor no está presente.

La razón de prevalencias puede ser un buen estimador de la razón de riesgos, o ***riesgo re-***

Tabla 14.2 Resultados de un estudio de corte transversal

		Enfermedad prevalente		
		Presente (E)	Ausente ($\bar{E}$)	
Factor	Presente (F)	a	b	a+b
	Ausente ($\bar{F}$)	c	d	c+d
		a+c	b+d	a+b+c+d

Donde:

a	Individuos con enfermedad prevalente y con factor presente
b	Individuos sin enfermedad prevalente y con factor presente
c	Individuos con enfermedad prevalente y ausencia de factor
d	Individuos sin enfermedad prevalente y ausencia de factor
a+c	Individuos enfermos
b+d	Individuos no enfermos
a+b	Individuos con factor presente
c+d	Individuos con ausencia del factor
a+b+c+d	Total de individuos valorados en el estudio
(a+c)/(a+b+c+d)	Prevalencia de la enfermedad en la población estudiada
(a+b)/(a+b+c+d)	Prevalencia de factor en la población estudiada
a/(a+b)	Prevalencia de enfermedad en personas con factor
c/(c+d)	Prevalencia de enfermedad en personas sin factor

lativo (RR) si se cumplen simultáneamente dos condiciones: la primera, que la duración de la enfermedad sea igual entre pacientes expuestos y no expuestos; y la segunda, que no haya habido migración hacia dentro o hacia fuera de estos grupos previamente a la puesta en marcha del estudio. Los estudios de corte transversal en sí mismos no permiten determinar el cumplimiento de las mencionadas condiciones, hecho que limita la interpretación de las asociaciones encontradas.

Por otro lado, la ***razón de suertes*** (RS) (OR) se calcula dividiendo la razón de factor en los enfermos sobre la razón de factor en los pacientes no enfermos [(a/c)/(b/d)]; o, de igual forma, dividiendo la razón de enfermedad entre las personas en quienes el factor está presente sobre la razón de enfermedad en las personas en quienes el factor no está presente [(a/b)/(c/d)]. Matemáticamente, ello corresponde a un producto cruzado, dado que al reducir cualquiera de las dos fórmulas anteriores obtendremos [(a × d)/(b × c)].

Corresponde ello al exceso o al defecto de ventaja (suertes, momios o, en la literatura anglosajona, *odds*) que tienen los individuos en quienes el factor está presente de tener la enfermedad o la condición frente a no tenerla, respecto a la misma ventaja en los individuos no expuestos.

Aunque el análisis estadístico de los estudios transversales y el de los de casos y controles presentan similitudes en el cálculo de los OR, tales estudios difieren en cuanto al diseño, y, por tanto, la interpretación de medidas, de sesgos y de aproximación a la causalidad son diferentes.

Al seleccionar a todos los individuos de una población o una muestra de ellos con pocos criterios de inclusión y de exclusión que permitan, en primera instancia, calcular las prevalencias, el control de variables de confusión debe hacerse en el análisis de los datos, y no por el diseño. Para ajustar estos efectos se hacen análisis estratificados o se emplean técnicas de análisis multivariado con regresión lineal múltiple o regresión logística.

La discusión sobre el uso de diversas técnicas estadísticas de acuerdo con las escalas de medición empleadas para las variables seleccionadas, las pruebas de bondad de ajuste y el control del efecto de diversas variables en diversos modelos sobrepasan el propósito de este capítulo, por lo cual se recomienda revisarlo en los capítulos correspondientes de análisis o en textos de bioestadística.

Ventajas

Más allá de que permiten determinar la frecuencia y la magnitud de condiciones, enfermedades o diversos factores, la mayor ventaja de los estudios de corte transversal, tal vez sea que permiten explorar, de forma eficiente, asociaciones entre múltiples condiciones o enfermedades y múltiples factores, no considerados *a priori*.

Las encuestas poblacionales permiten recolectar simultáneamente una gran cantidad de datos demográficos, clínicos y epidemiológicos, a partir de los cuales pueden explorarse asociaciones. A la vez, estas aportarán elementos que permiten generar nuevas hipótesis de investigación, la cuales, a su vez, serán estudiadas con diseños de mayor jerarquía.

Las asociaciones encontradas pueden contribuir a identificar factores de riesgo, siempre y cuando las exposiciones exploradas no se modifiquen en el tiempo. Sin embargo, como ya se mencionó, algunos de los factores asociados hallados en dichos estudios no solo no son factores de riesgo, sino, incluso, pueden ser consecuencia de la condición o la enfermedad.

Por lo general, esta clase de estudios se llevan a cabo en periodos cortos. Ello está especialmente ligado a los periodos de recolección de datos (regularmente, durante una única evaluación) y en un solo momento en el tiempo, y a la no necesidad de seguimiento de los individuos en la población o en la muestra estudiada. Tales características también hacen que puedan conducirse a menores costos, en comparación con otros diseños. Permiten, además, identificar las condiciones o las enfermedades más prevalentes, o bien, las de mayor duración en la población; así, brindan información útil para planear y administrar servicios de salud.

Limitaciones

Al valorar la presencia de la enfermedad o de la condición de salud al mismo tiempo que otros múltiples factores, los estudios de corte transversal no permiten identificar la secuencia de los eventos en el tiempo, y, por ende, tienen un aporte muy limitado para la inferencia de etiologías o de causalidades.

Además de lo anterior, se exponen a sesgos inherentes a la duración de las condiciones o las enfermedades objeto de observación, y por ello al hacerlos se puede sobreestimar o sobrerrepresentar los casos de larga duración y subrepresentar o subestimar los de corta duración. Cuando la condición, la enfermedad o factor asociado ocurren con baja frecuencia o son de corta duración, solo se captará información de un número reducido de las personas que los padecen.

En múltiples ocasiones se hace difícil interpretar los factores asociados en los estudios de corte transversal, pues tales factores pueden estar sobreestimados o, por el contrario, subestimados. Tal situación se hace especialmente relevante cuando la exposición cambia con el tiempo, como por ejemplo, cuando una persona deja de fumar o cambia de trabajo. A la vez, cuando una exposición se asocia a una enfermedad leve y de larga duración, la frecuencia de exposición será elevada en los casos. En cambio, si la exposición produce una alta letalidad de la enfermedad, entonces la frecuencia de exposición será muy baja entre los casos, y la asociación puede resultar negativa.

Cuando el propósito principal del estudio es determinar la prevalencia, puede ser que no tenga poder suficiente para hallar asociaciones significativas entre las condiciones o las enfermedades y los factores asociados.

Reporte y publicación de estudios de corte transversal

A menudo, la comunicación de estudios observacionales no es lo bastante clara y detallada para evaluar las fortalezas y las debilidades de la investigación. Para mejorar este aspecto, se han desarrollado algunas iniciativas, dentro de las cuales la más importante puede ser STROBE (por las siglas en inglés de Strengthening the Reporting of Observational Studies in Epidemiology). STROBE propone una lista de chequeo de 22 puntos, los cuales se presentan en las **tablas 14.3 y 14.4.**

Tabla 14.3 Ventajas y desventajas del diseño de corte transversal

Ventajas
Rapidez en la recolección de información.
Menores costos que en los estudios de corte longitudinal.
Utilidad para establecer prioridades en la planeación y el uso de recursos.
Utilidad para determinar asociaciones entre variables y generar hipótesis.
Utilidad para la detección de casos habituales.
Desventajas
Limitaciones para inferir relaciones de causalidad.
En condiciones fatales, solo permiten identificar a sobrevivientes.
Poca utilidad en casos de rara ocurrencia.
Poca utilidad en enfermedades de corta duración.
No permiten predecir eventos futuros.

Tabla 14.4 Declaración STROBE: lista de puntos que deben incluirse en los reportes de estudios de corte transversal

Sección	Tema	Punto	Recomendación
Preliminares	Título	1	**(a) Indique, en el título o en el resumen, el diseño del** estudio con un término habitual.
	Resumen		**(b) Proporcione en el resumen una sinopsis informativa** y equilibrada de lo que se ha hecho y lo que se ha encontrado.
Introducción	Contexto/fundamentos	2	**Explique las razones y el fundamento científico de la** investigación que se comunica.
	Objetivos	3	**Indique los objetivos específicos, incluyendo cualquier** hipótesis preespecificada.
Métodos	Diseño del estudio	4	**Presente al principio del documento los elementos clave del** diseño del estudio.
	Contexto	5	**Describa el marco, los lugares y las fechas relevantes,** incluyendo los periodos de reclutamiento, de exposición, de seguimiento y de recogida de datos.
	Participantes	6	**Proporcione los criterios de elegibilidad, así como las** fuentes y el método de selección de los participantes.
	Variables	7	**Defina claramente todos los desenlaces, las exposiciones,** y las variables predictoras, confusoras y modificadoras del efecto. Si procede, proporcione los criterios diagnósticos.
	Fuentes de datos/medidas	8*	Para cada variable de interés, indique las fuentes de datos y los detalles de los métodos de valoración (medida). Si hay más de un grupo, especifique la comparabilidad de los procesos de medición.
	Sesgos	9	**Describa todas las medidas adoptadas para afrontar** posibles fuentes de sesgo.
	Tamaño muestral	10	**Explique cómo se determinó el tamaño muestral.**
	Variables cuantitativas	11	**Explique cómo se trataron las variables cuantitativas en el** análisis. Si procede, explique qué grupos se definieron, y por qué.
	Métodos estadísticos	12	(a) Describa todos los métodos estadísticos, incluidos los empleados para controlar los factores de confusión.
			(b) Describa todos los métodos utilizados para analizar subgrupos e interacciones.
			(c) Explique el tratamiento de los datos ausentes.
			(d) Si procede, describa los métodos analíticos empleados en la estrategia de muestreo.
			(e) Describa los análisis de sensibilidad.
Resultados	**Participantes**	**13***	(a) Indique el número de participantes en cada fase del estudio; por ejemplo, el número de participantes elegibles, analizados para ser incluidos, confirmados elegibles, incluidos en el estudio, los que tuvieron un seguimiento completo y los analizados.
			(b) Describa las razones para no participar en cada fase.
			(c) Considere el uso de un diagrama de flujo.

Continúa

Continuación

Sección	Tema	Punto	Recomendación
Resultados	Datos descriptivos	14*	(a) Describa las características de los participantes en el estudio (por ejemplo, demográficas, clínicas, sociales) y la información sobre las exposiciones y los posibles factores de confusión.
			(b) Indique el número de participantes con datos ausentes en cada variable de interés.
	Datos de las variables de resultado	**15***	**Indique el número de eventos resultado o medidas resumen.**
	Resultados principales	16	(a) Proporcione estimaciones no ajustadas, y, si procede, ajustadas por factores de confusión, así como su precisión (por ejemplo, intervalos de confianza del 95 %). Especifique los factores de confusión por los que se ajusta, y las razones para incluirlos.
			(b) Si categoriza variables continuas, describa los límites de los intervalos.
			(c) Si es pertinente, considere acompañar las estimaciones del riesgo relativo con estimaciones del riesgo absoluto para un periodo relevante.
	Otros análisis	17	**Describa otros análisis efectuados (por ejemplo, análisis de** subgrupos e interacciones, y de sensibilidad).
Discusión	Resultados clave	18	**Resuma los resultados principales con referencia a los** objetivos del estudio.
	Limitaciones	19	Discuta las limitaciones del estudio, teniendo en cuenta posibles fuentes de sesgo o de imprecisión. Discuta tanto sobre la dirección como sobre la magnitud de cualquier posible sesgo.
	Interpretación	20	Proporcione una interpretación global y prudente de los resultados; considere: objetivos, limitaciones, multiplicidad de análisis, resultados de estudios similares y otras evidencias relevantes.
	Generabilidad	21	**Discuta la posibilidad de generalizar los resultados (validez** externa) del estudio.
Otra información	**Financiación**	**22**	**Especifique la fuente de financiación y el papel de los** patrocinadores del estudio, y, si procede, el del estudio previo en el que se basa su artículo.

* Dé información separada para los grupos de expuestos y no expuestos.

Nota: Un artículo de explicación y elaboración discute cada ítem de la lista de chequeo, y brinda una perspectiva metodológica y ejemplos publicados de reportes transparentes. La lista de chequeo STROBE debe usarse conjuntamente con este artículo, que está disponible de forma gratuita en los sitios web de PLoS Medicine (www.plosmedicine.org/), Annals of Internal Medicine (http://www.annals.org/) y Epidemiology (http://www.epidem.com/). En la página web de STROBE (http://www.strobe-statement.org) aparecen las diferentes versiones de la lista, correspondientes a los estudios de cohortes, a los estudios de casos y controles y a los estudios transversales.

Bibliografía

Bhattacharya K. Introduction to statistics for medical students. Oxford: University of Oxford; 2005.

Hernández B, Velasco-Mondragón HE. Encuestas transversales. Salud Púb México. 2000;42:447-55.

Levin KA. Study design III: Cross-sectional studies. Evid Based Dent. 2006;7:24-5.

Lietz P. Cross-sectional research methods. En: Husen T, Postlethwaite TN, editores. *The international encyclopedia of education. 2nd ed.* Oxford, UK: Pergamon; 1994. p. 1213-20.

Olsen C, George DM. Cross-sectional study design and data analysis. The Young Epidemiology Scholars Program (YES). Estados Unidos: YES; 2004.

Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

Schiaffino A, Rodríguez M, Pasarín MI, et al. Odds ratio o razón de proporciones? Su utilización en estudios transversales. Gac Sanit. 2003;17:70-4.

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344-9. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Bull World Health Organ. 2007;85:867-72.

CAPÍTULO

15 El experimento clínico

JUAN MANUEL LOZANO LEÓN • JUAN GABRIEL RUIZ PELÁEZ

INTRODUCCIÓN

La comparación es una de las bases del método epidemiológico, y, sin duda, la piedra angular del experimento clínico y considerada actualmente el mejor método para determinar si una intervención en salud produce más beneficio que daño. En este capítulo se discuten los principales aspectos de los fundamentos teóricos y del diseño de tal tipo de estudios.

DEFINICIÓN

Aunque hay múltiples clasificaciones de los diseños de investigación clínica o epidemiológica, la mayoría de los autores está de acuerdo en la existencia de dos grandes tipos de estudios: 1) experimentales y 2) no experimentales. La diferencia básica estriba en el método empleado para asignar a los sujetos del estudio a la exposición o a la intervención de interés[1-4]. En los estudios experimentales, el investigador es quien determina cuáles sujetos recibirán una u otra exposición[1].

El objetivo del experimento es determinar si las modificaciones en la variable ***independiente*** (la que se encuentra bajo el control del observador, y que corresponde a la intervención que se estudia) cambian algún desenlace (variable *dependiente*, o variable desenlace).

EJEMPLO

Para evaluar un inhibidor selectivo de la ciclooxigenasa 2 (ICOX-2) y compararlo con un antiinflamatorio no esteroideo (AINE) convencional, el investigador asigna a un grupo de sujetos a recibir ICOX-2, y a otro grupo, a recibir AINE (*variables independientes*). En este caso, se trata de un experimento, pues la exposición a uno u otro medicamento depende exclusivamente del investigador.

En contraste, en los estudios no experimentales el investigador no asigna la variable independiente, sino que observa los efectos de un "experimento" natural en sujetos que, voluntaria o involuntariamente, se han sometido a una exposición, y los compara con quienes no lo han hecho.

[1] En la mayor parte de los casos, estas exposiciones son un tratamiento nuevo o experimental y otro usual o de control. A lo largo de este capítulo se usará el término ***intervención*** más que el de ***tratamiento***, pues la exposición puede ser a un programa educativo, a una vacuna, a un procedimiento quirúrgico o a un medicamento.

EJEMPLO

La evaluación del efecto que pueda tener la exposición al humo de leña en la producción de enfermedad pulmonar debe hacerse comparando la frecuencia de enfermedad pulmonar en quienes se expusieron a lo largo de su vida al humo de leña con la de quienes no lo hicieron, o comparando la frecuencia de exposición en quienes tienen enfermedad pulmonar con la frecuencia en quienes no la padecen. *En ninguno de los dos casos el investigador tuvo influencia en la exposición*.

Es necesario aclarar que para la mayoría de autores un estudio experimental es el estudio donde la asignación de los sujetos a sus respectivos grupos se hace mediante el azar. Cualquier otra forma de asignación corresponde a un pseudoexperimento.

El propósito central de un experimento clínico es comparar los resultados de una exposición en un conjunto de sujetos con los de otro grupo, sometido a una exposición de control. Un aspecto muy importante de dicho estudio es obtener grupos tan similares como sea posible, en aspectos diferentes de la exposición, pero que puedan influir sobre el resultado (también llamados ***variables modificadoras del efecto***, ***variables pronósticas*** o ***variables de confusión***), de forma que la única diferencia entre los grupos sea la exposición que reciben. Si se logra esta comparabilidad, el hallazgo de diferencias en la frecuencia del desenlace (variable dependiente) solo podrá atribuirse a las intervenciones, pues en los otros aspectos los grupos eran semejantes.

Solo cuando el desenlace de una condición clínica es previsible de forma inequívoca, se puede evaluar la utilidad de una intervención clínica sin contar con controles[5]. Por ejemplo, la eficacia de la penicilina en el tratamiento de pacientes con neumonía —de quienes, según la experiencia previa, se esperaba que murieran por la enfermedad— fue demostrada por medio de series de casos, donde todos los sujetos fueron tratados con el antibiótico[6]. Los dramáticos resultados, que no podían atribuirse a ningún otro factor sino al empleo del medicamento, llevaron a la aceptación de los antibióticos como agentes terapéuticos en la enfermedad, y volvieron antiético (y, probablemente, inútil) hacer un experimento clínico.

No obstante lo anterior, la investigación clínica actual enfrenta desafíos diferentes. En muchos casos, se pretende demostrar que las nuevas intervenciones mejoran los resultados que ya han sido obtenidos con las terapias convencionales. Como la diferencia esperada es mucho más pequeña que la observada cuando comenzó el uso de la penicilina, es necesario refinar los métodos de estudio empleando técnicas que permitan aislar los efectos de las intervenciones de las muchas otras variables capaces de influir sobre el desenlace de las enfermedades.

Además del tamaño de la diferencia esperada, hay otras razones que obligan a emplear controles al evaluar una intervención. En primer lugar, se sabe que los participantes en un estudio pueden modificar su comportamiento al saberse observados (usualmente, mostrando una respuesta positiva o benéfica). Este fenómeno, llamado *efecto Hawthorne*, fue descrito en un estudio que evaluó los efectos de la iluminación sobre la productividad de los empleados en una planta eléctrica[7]. Los investigadores observaron cómo al mejorar la iluminación aumentó la productividad, y que al regresar a los niveles iniciales de iluminación, la productividad se redujo, pero se estabilizó en niveles superiores a los hallados antes de la intervención. La explicación para tal fenómeno fue el cambio en el comportamiento de los trabajadores al saberse observados.

En segundo lugar, si se selecciona un grupo de sujetos porque presentan valores extremos en una distribución (por ejemplo, presión arterial o colesterol elevados) puede esperarse que en las mediciones subsecuentes de dicho fenómeno se obtengan valores menos extremos. Es lo que se conoce como ***regresión al promedio***, y tiene razones puramente estadísticas (***variación aleatoria***)[8]. La selección de personas que tienen un valor inicial por encima de un punto de corte puede incluir a algunos sujetos que en ese momento tenían cifras inusualmente altas, y que en mediciones posteriores muestren cifras más bajas. Dichos resultados más bajos tienden a arrastrar hacia abajo el promedio del grupo. Si ello coincide con una intervención, el investigador puede atribuir, erróneamente, la "mejoría" del grupo a esta regresión al promedio.

En tercer lugar, una intervención puede tener efectos (usualmente, benéficos) en quien la recibe, aun si no hay un mecanismo de acción que explique tal efecto. Conocido como ***efecto***

placebo, esta reacción puede llegar a ser tan común (como el 40 %) en casos especiales como en dolor. La administración de un placebo (sustancia inerte) puede lograr mejoría en muchos aspectos clínicos, y debe ser tenida en cuenta, pues bien podría explicar un alto porcentaje de la respuesta que se observe.

Por último, a diferencia de lo descrito para el primer empleo de la penicilina en el tratamiento de la neumonía, los efectos de muchas intervenciones terapéuticas no son fácilmente predecibles. Todos los clínicos se han enfrentado a una situación en la cual piensan que un paciente habría presentado el mismo resultado con o sin un tratamiento determinado.

La presencia de un grupo control en un experimento le permite al investigador neutralizar el impacto que puedan tener los efectos descritos (placebo, Hawthorne y regresión al promedio), pues se presentan por igual en el grupo de la intervención y en el grupo de control. De esta manera, la diferencia observada puede atribuirse a diferencias entre las intervenciones, y puede descartarse la intervención de los efectos mencionados.

TIPOS DE EXPERIMENTOS CLÍNICOS

Experimento clínico aleatorizado ( ) convencional

El objetivo de la mayoría de los experimentos es demostrar que la nueva intervención produce mejores resultados que la convencional. En el experimento clínico convencional, los sujetos elegibles son distribuidos en dos grupos independientes, llamados ***experimental*** y ***control***, cada uno de los cuales recibe la intervención correspondiente.

El seguimiento de los participantes permite medir los resultados en cada grupo y compararlos mediante el uso de las técnicas estadísticas apropiadas, para establecer la dirección y el tamaño de las diferencias, así como la probabilidad de que las diferencias observadas, si existen, hayan sido producidas por el azar. Si dicha probabilidad es muy baja, el investigador concluye que los dos tratamientos producen resultados diferentes. La mayor parte del presente artículo estará dedicada a los principios del diseño de este experimento convencional.

EJEMPLO

A un investigador le interesa establecer si la administración de sacarosa a neonatos durante procedimientos dolorosos menores (como punciones venosas) reduce el dolor, en comparación con lo observado al administrarles agua[2]. La **figura 15.1** ilustra cómo puede contestarse dicha pregunta mediante un experimento clínico convencional. A un grupo de neonatos se lo asigna a recibir la solución de sacarosa, y otro grupo es asignado a recibir agua, inmediatamente *antes* de la punción venosa. Luego, el investigador determina el resultado en cada grupo estableciendo, por ejemplo, la proporción de sujetos que lloran, la duración y la intensidad del llanto o cualquier otro desenlace apropiado.

Experimento cruzado

Este diseño fundamental puede tener variaciones. En el experimento cruzado (***cross-over***), cada participante recibe una intervención inicialmente (bien sea la experimental, o bien, la de control) por un tiempo. Luego del periodo inicial, se suspenden las intervenciones en los dos grupos, se espera un tiempo para cuando desaparezcan sus efectos y se cruzan los tratamientos: el grupo que había recibido la intervención experimental recibe ahora la convencional, y viceversa[9-11].

En este diseño, todos los participantes reciben las dos intervenciones de interés, y lo que se asigna es su orden. Como cada sujeto es su propio control, la medición de la variable dependiente (desenlace o resultado) presenta menos variabilidad que al hacerla en sujetos diferentes, lo cual reduce de manera significativa el número de personas requeridas en el experimento.

El diseño cruzado permite no solo reducir la variabilidad que se presentaría en un experimento clínico convencional, sino también, reducir el número de sujetos necesarios para el estudio, pues cada participante contribuye al menos con dos periodos de observación.

Por otra parte, este diseño exige que al suspender el primer tratamiento la condición del sujeto sea igual a la existente antes de recibirla, pues si la intervención inicial cura o mejora significativamente la condición, el segundo tratamiento se aplicaría a un individuo distinto del que recibió la primera. Ello limita el uso de dicho diseño a condiciones no curables, de curso crónico y que permanezcan estables en el tiempo.

[2] Este ejemplo fue sugerido por Luis Gabriel Cuervo, coautor en ediciones anteriores de este libro.

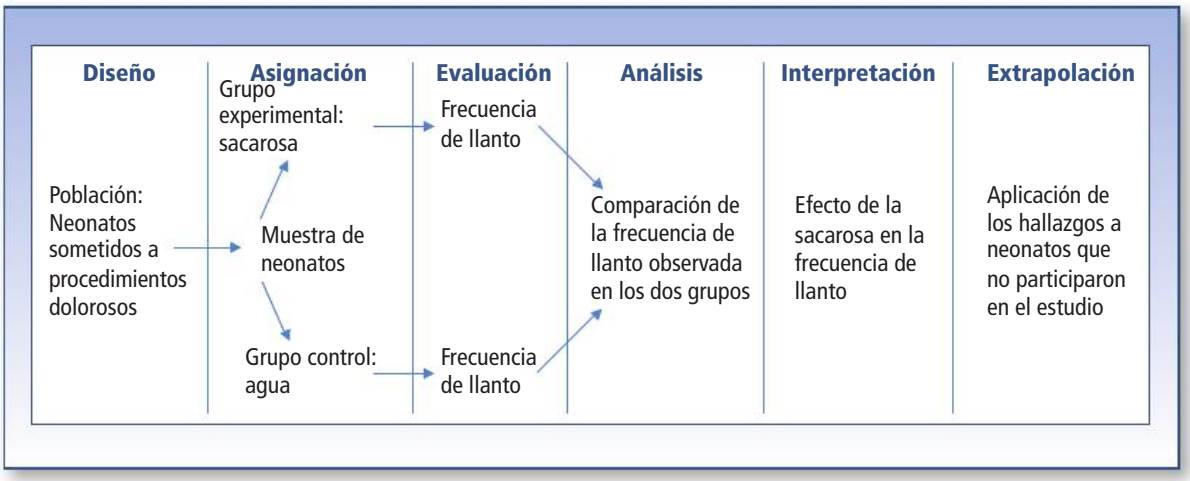


Figura 15.1 ECA para comparar sacarosa y agua en procedimientos dolorosos en neonatos.
Fuente: Elaboración propia.

EJEMPLO

Puede utilizarse el diseño cruzado en la migraña clásica en sujetos que no tengan variaciones periódicas reconocidas, en enfermedades dermatológicas crónicas y con poca variabilidad (pitiriasis rosada, psoriasis) o en condiciones como asma, artritis reumatoidea o colon irritable, en sujetos con enfermedad estable y en quienes no se estén evaluando medicamentos modificadores de la enfermedad (no podría evaluarse con este diseño el efecto de metotrexate en la artritis reumatoidea, por su acción modificadora de la enfermedad; pero sí podría compararse un AINE con un ICOX-2 en una población que reciba, en su totalidad, metotrexate).

Adicionalmente, cuando el efecto de la intervención es prolongado debe existir un "periodo de lavado", durante el cual el participante no reciba ninguna intervención, lo cual garantiza que la evaluación de la segunda intervención no esté contaminada con el efecto de la primera.

EJEMPLO

Puede contestarse la pregunta sobre el efecto analgésico de la sacarosa mediante un diseño cruzado con neonatos que sean sometidos a varias punciones. Un grupo será asignado a recibir primero la sacarosa, y a la siguiente vez, el agua; en tanto, el otro grupo será asignado a recibir la primera vez agua, y la segunda vez, la solución con sacarosa. La comparación de los desenlaces con cada una de las intervenciones permitirá comparar la efectividad de cada una de las intervenciones.

Experimento clínico de equivalencia

Cuando se considera que la nueva intervención puede tener algunas ventajas sobre el tratamiento convencional (menor costo, más comodidad o menos efectos secundarios), el investigador puede estar interesado en demostrar que el efecto de los dos tratamientos sobre la variable principal es similar. Si ello es así, las ventajas de la nueva intervención la harían el tratamiento de elección. En este caso, se conduce un experimento de equivalencia, que tiene su principal aplicación cuando el tratamiento convencional produce resultados satisfactorios (en caso de no ser así, el objetivo debería ser el desarrollo de nuevos y mejores tratamientos[12,13].

Si el investigador quiere demostrar equivalencia, debe plantear su pregunta de investigación de la siguiente forma: si se pretendía demostrar que la nueva intervención era mejor que la convencional y el estudio no demostró dichas diferencias,

EJEMPLO

Si se asume que se ha demostrado que la sacarosa es mejor que el agua para reducir el dolor en neonatos sometidos a procedimientos dolorosos, un investigador puede estar interesado en determinar si se obtienen resultados equivalentes reemplazando la sacarosa por una solución de dextrosa, la cual tiene como ventaja que ya está disponible rutinariamente en preparación estéril en cualquier institución de salud. Si se demuestra que las dos soluciones son equivalentes en cuanto a reducción de dolor, se seleccionaría la solución de dextrosa, por la ventaja mencionada.

es incorrecto suponer que los tratamientos son equivalentes[14,15]; sobre todo, porque tal hallazgo puede deberse a la falta de poder estadístico para demostrar las diferencias.

Experimento clínico de no inferioridad

Se trata de una riesgosa variación del experimento de equivalencia, en la que se propone, para la comparación entre dos intervenciones, una diferencia en el resultado, que va a ser considerada "no significativa", por lo cual puede concluirse que, aun si existe esa diferencia, pues se la definió como no significativa, el tratamiento evaluado es "no inferior" al otro. Podría definirse, por ejemplo, que al comparar dos intervenciones para reducir el riesgo de mortalidad, se pueden aceptar diferencias donde el tratamiento de interés sea hasta el 10 % peor que el tratamiento estándar.

Claramente, hay un riesgo en esta aproximación: la subjetividad del límite lo hace claramente susceptible de sesgos. Un riesgo adicional es que el enfoque de no inferioridad implica que para el análisis estadístico se considerarán solo los resultados peores (en lenguaje estadístico, solo se hará análisis a una cola de la distribución, en contraposición con el análisis tradicional, que es a dos colas), y, es muy claro, la posibilidad de obtener resultados estadísticamente significativos sin una verdadera diferencia (***error tipo I*, o de falso positivo) aumenta.**

Experimento clínico factorial

Ocasionalmente, el investigador puede evaluar dos intervenciones distintas, para condiciones también muy diferentes, utilizando la misma población mediante un experimento factorial. Dadas las dificultades y los costos de los experimentos clínicos, es atractivo utilizar el mismo grupo de individuos para dos experimentos simultáneos. Tal diseño debe ser utilizado solo cuando haya garantía de que el efecto de cada una de las intervenciones es el mismo en presencia o en ausencia de la otra intervención; es decir, cuando las intervenciones son independientes y también lo son las condiciones estudiadas en cada uno de los grupos. Como cabe esperarlo, dicho tipo de estudio es más complejo, y requiere considerar con cuidado los efectos del empleo simultáneo de varios medicamentos[17].

EJEMPLO

En un experimento clínico factorial se evaluó, de manera simultánea, la efectividad de agregar prednisolona o inmunoterapia con *Mycobacterium indicus pranii* con el manejo con antimicrobianos y antirretrovirales en pacientes con pericarditis tuberculosa[16]. En tal estudio hubo cuatro grupos de tratamiento: 1) prednisolona e inmunoterapia; 2) prednisolona y placebo de inmunoterapia; 3) placebo de prednisolona e inmunoterapia, y 4) placebos tanto de prednisolona como de inmunoterapia.

Experimento de asignación por grupos

El diseño de asignación por grupos **se emplea** cuando hay alta probabilidad de que los sujetos asignados a un tratamiento puedan recibir también la intervención del otro grupo (***contaminación*)[18]. Asignar toda la población al mismo** grupo reduce esta posibilidad. Sin embargo, la asignación por grupos implica aumentar el tamaño de la muestra cuando se compara con la necesaria para el diseño tradicional.

EJEMPLO

Un estudio evaluó la efectividad de complementar el lavado de las manos con antisépticos que contengan alcohol con las prácticas habituales de lavado de las manos, para reducir la frecuencia de enfermedad diarreica y de infección respiratoria en niños entre 1 y 5 años de edad que asistían a guarderías comunitarias con limitado acceso a agua corriente en Bogotá. Dada la pregunta del estudio, se consideró impráctico asignar a algunos de los niños de una guardería a recibir cuidado luego de que los cuidadores les lavaran sus manos con los métodos habituales, y a otros, a recibir atención por los mismos cuidadores luego de complementar el lavado de las manos con antisépticos. El experimento asignó de forma aleatoria 21 centros (con 794 niños) a usar antisépticos, y 21 centros (933 niños), al lavado tradicional[19].

Experimento clínico con análisis secuencial

Otra variante del experimento clínico es con una modalidad particular de evaluación de resultados, llamado ***análisis secuencial*. El principal** atractivo de tal tipo de diseño es que no tiene

un tamaño de muestra fijo, por lo cual puede llegar a obtener un resultado estadísticamente significativo con el menor número posible de sujetos necesarios para encontrar una diferencia previamente establecida.

Los sujetos ingresan al estudio en parejas, y cada uno es asignado a recibir una de las dos intervenciones. Mediante el seguimiento, se determina si el tratamiento tiene éxito o no en cada miembro del par, lo que limita los hallazgos a solo cuatro posibilidades: 1) los dos presentan un desenlace exitoso, 2) los dos fracasan, 3) el experimental tiene éxito y 4) el control no tiene éxito, o el control es exitoso, y el experimental, no. Las dos primeras opciones corresponden a empates y se descartan; solo los resultados de parejas no empatadas se consideran en el análisis.

Estos resultados no empatados se van acumulando hasta cuando una de las dos opciones (parejas en las que fue más exitoso el tratamiento convencional, o parejas en las que fue más exitoso el tratamiento experimental) es más común que la otra y alcanza una cifra previamente establecida, con criterio estadístico para mostrar significación. En ese momento termina el estudio. Si ninguno de los dos tratamientos es definitivamente superior, el diseño también prevé el cálculo de un límite de parejas estudiadas compatible con el hecho de que no haya diferencias entre los tratamientos.

Las principales limitaciones de este diseño son que solo permite evaluar dos tratamientos, que la variable dependiente ***tiene que ser*** dicotómica (éxito/fracaso, vivo/muerto) y que el desenlace debe poder determinarse muy pronto después de iniciada la intervención. Adicionalmente, los métodos estadísticos necesarios para el análisis de los resultados son bastante complejos[16,20].

EJEMPLO

En un estudio factorial se evaluó la efectividad del suplemento con ácido fólico o con otras vitaminas para prevenir defectos del tubo neural en nuevas gestaciones de 1817 mujeres con antecedente de un hijo con dicha anomalía. Diseñado desde el comienzo para evaluación de resultados con la técnica de análisis secuencial, el estudio pudo ser detenido cuando se alcanzaron 27 casos de defectos neurales: 6 en el grupo de ácido fólico y 21 en los restantes grupos; o sea, hubo una reducción del 72 % en el riesgo de aparición del defecto[21].

Experimento de descontinuación

En el experimento de descontinuación se parte de un grupo de individuos que ya estén recibiendo la intervención cuya efectividad se desea evaluar. En los que tienen una respuesta calificada como aparentemente favorable, se inicia el estudio mediante aleatorización a una de dos opciones: o bien continuar la terapia que ya mostró efectividad aparente, o bien, a recibir una intervención de control (que puede ser placebo, según el caso)[16,22].

Este tipo de estudio es particularmente útil para evaluar el efecto de terapias no curativas de larga duración; en especial, si el efecto esperado es de pequeña magnitud, así como cuando, por razones éticas o de factibilidad, se quiere limitar el empleo del tratamiento control.

Este diseño tiene algunas limitaciones. La más importante es que no permite evaluar intervenciones que produzcan resultados definitivos: por ejemplo, procedimientos quirúrgicos. Adicionalmente, la selección de individuos que presenten una respuesta en apariencia satisfactoria implica que se descarta a los poco adherentes o a quienes han suspendido la medicación por haber experimentado efectos adversos, lo cual, a su vez, puede llevar a sobreestimar los efectos benéficos, a subestimar los riesgos y la toxicidad y, en general, a una limitada generalizabilidad de los resultados.

EJEMPLO

El Ensayo de Supresión de Arritmias Cardíacas (CAST, por las siglas en inglés de *Cardiac Arrhytima Supression Trial*) se inició con 1498 pacientes con antecedente de infarto del miocardio, en quienes se había logrado una reducción del 80 % en la frecuencia de extrasístoles ventriculares mediante el uso de encainide o de flecainide. Dichos sujetos fueron asignados a continuar el antiarrítmico o a recibir placebo. El estudio debió ser terminado de manera prematura, al observarse un exceso de muertes (la mayor parte de ellas, por condiciones cardíacas) en el grupo que siguió recibiendo el antiarrítmico, en comparación con el de placebo (63 *vs*. 26, respectivamente)[23].

Megaexperimentos clínicos

La búsqueda de efectos de tamaño reducido, pero con importancia clínica o epidemiológica,

ha llevado a realizar grandes experimentos clínicos, en los que se comparan los resultados de las intervenciones de interés en grupos de varios miles de sujetos. Estos megaexperimentos suelen, además, tener características que buscan hacerlos tan simples como sea posible, lo cual incluye pocos criterios de inclusión y de exclusión, criterios de inclusión amplios, intervenciones simples y desenlaces fáciles de medir[24].

EJEMPLO

Un grupo internacional (que incluyó a investigadores de 4 centros en Colombia, así como de Argentina y de Cuba) condujo un ECA en el que se asignaron gestantes con preeclampsia a recibir sulfato de magnesio (5068 mujeres) o placebo (5068 mujeres). El desenlace principal fue la *frecuencia de eclampsia*; es decir, la aparición de convulsiones. Otros desenlaces incluyeron la muerte por cualquier causa en la gestante o en su hijo y la admisión de los neonatos a cuidado intensivo neonatal, entre varios más[25].

Otro ejemplo evaluó una hipótesis práctica: la administración de una dosis fija de un hipolipemiante (estatina) a sujetos con riesgo moderado o alto de enfermedad coronaria en quienes no hubiera dislipidemia evidente, ¿reducía el riesgo de mortalidad coronaria y de eventos vasculares cerebrales? El estudio Heart Protection Study (HPS) se llevó a cabo en 20 536 sujetos de varios centros, y mostró que la dosis fija de estatina en sujetos con riesgo moderado o alto y sin clara dislipidemia generaba beneficios claros[26].

Experimento clínico individual (N = 1)

Finalmente, en el experimento clínico individual (o ***N* = 1**) **se asigna solo a un sujeto a recibir,** secuencialmente y de manera aleatoria, una serie de tratamientos (en este aspecto, es similar al estudio cruzado). La evaluación de las variables dependientes permite establecer la eficacia de las intervenciones y determinar si alguna es superior[27,28]. **Este tipo de diseño tiene aplicaciones** limitadas en la investigación convencional, pero puede ser muy útil para determinar la terapia más efectiva para un caso individual.

También en este caso deben cumplirse las recomendaciones para los experimentos cruzados: solo puede hacerse en condiciones crónicas y en las que no haya modificación de la enfermedad, debe dejarse un periodo suficiente de lavado y debe garantizarse que la condición del sujeto es igual al comienzo de cada uno de los periodos de intervención, los cuales se asignan de manera aleatoria.

COMPONENTES DE UN EXPERIMENTO CLÍNICO CONTROLADO

La planeación del experimento debe ser tan cuidadosa como la de cualquier otro estudio. Debe empezar con el protocolo de estudio, que es el documento en el que se consignan la pregunta de investigación, el fundamento científico, los objetivos, el diseño y la organización metodológica y logística.

El protocolo debe estar terminado en detalle antes de la inclusión de sujetos en el experimento, y no debe sufrir modificaciones posteriores. Esta es la garantía que tiene el investigador acerca de la solidez metodológica de su estudio y la que ofrece a los potenciales usuarios de sus resultados. De hecho, el protocolo es imprescindible para obtener aprobación por parte de los comités de ética de los centros donde se va a realizar el estudio y para obtener financiación. La **tabla 15.1 presenta los principales componentes** de un experimento clínico. A continuación se describe con más detalle cada uno de estos puntos.

Diseño del estudio

La pregunta de investigación

Todo experimento clínico debe tener una pregunta primaria o principal, definida claramente y para la cual se planea el estudio. La pregunta debe tener cuatro componentes claros y explícitos: 1) la población que será estudiada (definición de la población a la cual será posible extrapolar los resultados), 2) la intervención experimental que se quiere evaluar, 3) la intervención de control y 4) las variables de respuesta o dependientes (*variables de desenlace*) **que se utilizarán para comparar** los resultados de las intervenciones[29].

En algunos casos, la pregunta puede plantearse como la prueba de una hipótesis, lo cual implica, además, postular la magnitud de la diferencia que se espera encontrar, en las variables dependientes, entre los grupos del experimento.

Tabla 15.1 Componentes del protocolo de un ECA

Diseño del estudio
• Pregunta del estudio • Población (criterios de inclusión y de exclusión) • Consideraciones generales sobre tamaño de muestra y poder estadístico • Asignación • Definición de grupos de comparación: aleatorización; estratificación • Administración de las intervenciones
Evaluación
• Evaluación de base • Evaluación de resultados: estrategias de seguimiento, estudio abierto o ciego • Evaluación de adherencia a las intervenciones
Análisis
• Comparación de base • Comparación de los resultados: análisis por intención de tratar o por protocolo • Ajuste por variables de confusión • Interpretación
Extrapolación y generalización de los hallazgos

Fuente: Elaboración propia.

Además de la pregunta principal, el **ECA** puede incluir preguntas secundarias[29]. Aunque en algunos casos el estudio se diseña para dar respuesta específica a varias preguntas secundarias, es más común que la información obtenida para la pregunta principal se utilice para dar respuesta a las preguntas secundarias.

Estas preguntas secundarias pueden ser de dos tipos. En el primero se trata de determinar el impacto de la intervención sobre variables diferentes de la variable dependiente principal; en el segundo tipo se busca establecer si la intervención produce resultados distintos en diferentes subgrupos de sujetos, conformados según la edad, el sexo, la duración del proceso, su gravedad, etc.

EJEMPLO

En el estudio de la administración de sacarosa a neonatos, el desenlace principal puede ser la diferencia en la proporción de sujetos que lloran, y, secundariamente, si hay diferencias significativas bien sea en la duración y en la intensidad del llanto, o bien, en la frecuencia de punciones venosas fallidas.

Las preguntas secundarias deben ser pocas, contar con bases científicas firmes y postularse claramente antes de iniciar el estudio, para evitar sospechas posteriores de que se haya hecho una ***expedición de pesca con los datos***, nombre que se da a la búsqueda, ***a posteriori***, de preguntas que puedan ser contestadas con los datos encontrados, procedimiento a todas luces inadecuado.

Si la información no se recolectó a propósito desde el comienzo para contestar las preguntas secundarias, la capacidad del estudio para darles respuestas definitivas es limitada, y la confiabilidad de los resultados puede no ser adecuada, pues no se programó la recolección de ***esos*** **datos específicamente, o no se vigiló** la obtención de los datos. Además, debe recordarse que a medida que aumenta el número de pruebas estadísticas realizadas a una base de datos aumenta la probabilidad de encontrar diferencias significativas, pero debidas al azar. Por ello, se considera que lo más adecuado es ver las respuestas a las preguntas secundarias como hipótesis para posteriores estudios. El análisis de subgrupos presenta múltiples dificultades, y debe ser considerado con cautela[30].

Población de estudio

Seleccionar la población de estudio que debe regirse por criterios de elegibilidad, basados en los beneficios potenciales, en la posibilidad de efectos adversos y en las condiciones que hagan poco deseable la participación del sujeto en el estudio[31]. Generalmente, la elegibilidad se plantea en términos de criterios de inclusión y de exclusión.

Criterios de inclusión

Permiten definir el perfil del sujeto que se quiere admitir en el estudio. Usualmente, corresponden a características clínicas (por ejemplo, los parámetros para diagnosticar una enfermedad y su duración), demográficas (edad y sexo) y de ubicación geográfica y temporal (las instituciones donde se adelantará el estudio y el periodo de ejecución de este). Tales criterios definen la población de la cual se podrá extraer la muestra que será estudiada.

EJEMPLO

El megaexperimento ya mencionado (HPS) buscaba amplia generalizabilidad, así que se planeó con pocos criterios de inclusión, que, además fueran amplios: podía incluirse a cualquier hombre o mujer entre los 40 y los 80 años si tenía niveles de colesterol total superiores a 135 mg/dl y con riesgo sustancial de enfermedad coronaria (definido como antecedente de enfermedad coronaria, vascular o cerebral, diabetes o hipertensión arterial).

Criterios de exclusión

Los criterios de exclusión, por su parte, son las características que obligan a excluir del estudio a los sujetos que cumplen los criterios de inclusión. Las principales razones para ello son de tipo médico (se excluye a quienes tengan insuficiencia renal si el medicamento tiene eliminación principalmente por esa vía, o se excluye a los diabéticos que tienen riesgo aumentado de infección y respuesta menor a la terapéutica en un estudio que compare antibióticos, etc.); también, de tipo ético (el sujeto no acepta participar o hay condiciones que implican riesgos excesivos, tales como embarazo, lactancia o historia de reacciones secundarias con alguna de las intervenciones del estudio), y, finalmente, por razones de conveniencia (individuos a quienes no se podrá seguir con facilidad porque viven fuera del área de conducción del estudio o por otras características, como alcoholismo odrogadicción, analfabetismo, retraso mental o enfermedades psiquiátricas que lo inhabiliten para dar información, por ejemplo).

Como los criterios de exclusión se aplican solamente a quienes han cumplido los criterios de inclusión, es incorrecto decir que se excluirá a los sujetos que no cumplan los criterios de inclusión. También es incorrecto proponer como criterios de exclusión lo contrario de los criterios de inclusión.

Si los criterios de inclusión han definido la población de la cual se puede extraer la muestra (bronquíticos crónicos con exacerbación infecciosa, personas con dislipidemia y riesgo cardiovascular aumentado), los criterios de exclusión definen la población a la cual se podrán extrapolar los resultados del estudio, pues la definen más precisamente.

EJEMPLO

El estudio HPS excluyó, de la amplia gama de sujetos que podían ser incluidos (dislipidemia en hombres o mujeres entre 40 y 80 años con riesgo cardiovascular aumentado) quienes tuvieran enfermedad hepática o muscular (por los posibles efectos adversos de la estatina), enfermedad renal avanzada (por la vía de eliminación del medicamento), a quienes estuvieran recibiendo alguno de los medicamentos contraindicados con las estatinas, a mujeres no suficientemente protegidas contra embarazo (por el riesgo de malformaciones) y a personas que no pudieran dar información confiable (enfermedad psiquiátrica, etc.) o a quienes estuvieran en riesgo alto de muerte durante el periodo de estudio (cáncer o insuficiencia cardíaca severa).

Los criterios de elegibilidad deben definirse de manera no ambigua y en términos operativos más que académicos, de modo que el investigador pueda establecer de manera clara si un sujeto cumple o no con ellos. Por ejemplo, una definición académica de la artritis reumatoide como "enfermedad crónica y multisistémica de etiología desconocida […] cuya alteración característica es una sinovitis inflamatoria persistente […]"[32] no es práctica, pues no establece las características que debe tener un sujeto para hacer el diagnóstico de la enfermedad. En tal sentido, serían más útiles los parámetros establecidos por el American College of Rheumatology, entre los cuales están el tener al menos una articulación con sinovitis clínica no explicada por otra enfermedad y una puntuación igual o superior a 6 en una escala que mide, entre otros aspectos, el tipo y número de articulaciones afectadas, la duración de los síntomas y la presencia de marcadores serológicos. El protocolo debe también incluir la justificación de los criterios de inclusión y de exclusión.

En algunas oportunidades, los criterios de selección incluyen los resultados de pruebas de laboratorio o de anatomía patológica, los cuales pueden tomar cierto tiempo; en consecuencia, el tratamiento asignado puede iniciarse en sujetos a quienes luego se reconoce como no elegibles. ¿Cómo puede manejarse tal situación? Las opciones posibles incluyen: 1) no permitir el ingreso de ningún sujeto hasta no confirmar plenamente los criterios de selección; 2) permitir la inclusión antes de confirmar la elegibilidad y el

retiro posterior del análisis para los sujetos no elegibles, o 3) permitir el ingreso sin confirmar, e incluir en el análisis a todos los individuos admitidos; incluso, a los no elegibles. No hay una respuesta satisfactoria aplicable a todas las situaciones, por lo que el investigador debe considerar las ventajas y las limitaciones de cada opción. En cualquier caso, se debe decidir de antemano qué conducta se adoptará, y consignarla claramente en el protocolo[33].

Más o menos a menudo, se establece en los experimentos un periodo de gracia entre la firma del consentimiento informado y la toma de datos clínicos y de laboratorio, y solo se pasa a la segunda fase, de administración del medicamento unos días después, cuando se haya confirmado plenamente el cumplimiento de los criterios de admisión. Este periodo (***run-in period***) tiene particular importancia en los estudios metabólicos, en los cuales deben cumplirse criterios de laboratorio que de otra manera invalidarían el resultado o lo harían imposible de analizar.

EJEMPLO

Un estudio sobre el uso de estatinas (medicamentos que reducen, principalmente, el colesterol de baja densidad [cLDL]) en pacientes con síndrome coronario agudo necesita, como criterio de admisión, que haya hipercolesterolemia. Como no puede saberse si hay hipercolesterolemia sino con el resultado del laboratorio, debe haber un periodo de gracia en el estudio para que se completen los criterios de admisión antes de hacer aleatorización. No podría hacerse el examen antes de iniciarse el estudio, por cuanto es necesaria la firma del consentimiento informado para tomar la muestra. Y si se aleatoriza al sujeto antes de conocer los resultados, puede ocurrir que no haya hipercolesterolemia (y en ese caso no hay justificación para el tratamiento), o puede ocurrir que haya una severa hipertrigliceridemia (la cual, por razones técnicas, impide la determinación del cLDL). En este caso, un periodo de al menos dos días entre la firma del consentimiento y el resultado de la prueba de laboratorio permite el cumplimiento de los criterios y puede ser aprovechado, por ejemplo, para asegurar el cumplimiento de otras intervenciones en los dos grupos, tales como la dieta.

Definir la población del experimento determina varios aspectos importantes de este. En primer término, mientras mayor sea la homogeneidad respecto a otros determinantes del resultado diferentes de las intervenciones, y cuanto mayor sea el riesgo de presentar el desenlace, mayor será la eficiencia del estudio, lo que, desde el punto de vista estadístico, se traduce en un menor tamaño de la muestra.

En segundo lugar, y de forma estricta, los resultados del estudio solo se pueden generalizar a sujetos similares a los incluidos en él. Obviamente, dicho aspecto tiene cierto margen de interpretación, y puede ser evaluado más claramente por el usuario de los resultados que por el investigador. Sin embargo, si la población tiene características muy particulares, la generalización del estudio puede ser reducida. A menudo, hay cierta tensión entre ambos elementos: hacer más homogénea la población aumenta la eficiencia del estudio, pero puede reducir la aplicabilidad de sus resultados. En tales casos, los megaestudios sugieren mantener una amplia generalizabilidad, aun a costa de la reducción en la eficiencia; sin embargo, también para dicho aspecto tienen una respuesta los megaestudios, pues proponen la colaboración y la realización de estudios multicéntricos, que facilitan la recolección del mayor tamaño de muestra necesario.

Por último, los criterios de elegibilidad le permiten al lector conocer las características de la población estudiada. Si la población no se halla claramente definida, los usuarios potenciales de la información no sabrán si los sujetos que ellos ven en su práctica clínica se parecen o no a los del estudio, y, por lo tanto, no podrán determinar si se pueden beneficiar de los resultados. De hecho, esta es una de las razones para conducir megaexperimentos, pues el uso de criterios de selección amplios para permitir la inclusión de miles de sujetos en múltiples centros participantes también aumenta la generalización potencial de los hallazgos.

Una duda general entre los clínicos es si los estudios realizados en otros países (frecuentemente, más desarrollados) pueden generalizar sus resultados a los países en desarrollo. En contra de la muy común opinión, puede afirmarse que los resultados son generalizables (si los sujetos cumplen los criterios de admisión) hasta tanto no se demuestre lo contrario. Solamente se concluirá que no son generalizables los resultados en casos donde, si bien la población que podría recibir los beneficios del estudio cumple los criterios de

admisión, hay evidencia clara de diferencias significativas genéticas, étnicas, ambientales o culturales.

EJEMPLO

Un estudio en pacientes con múltiples factores de riesgo cardiovascular, llevado a cabo en Noruega, mostró que una intervención consistente en dieta con un alto contenido de pescado mejoraba significativamente el perfil lipídico, sin necesidad de medicamentos. Si bien los pacientes dislipidémicos en América Latina cumplen todos los criterios de admisión, la intervención es muy específica de Noruega, y no puede ser generalizable, pues la utilidad de la dieta no estriba en el alto consumo de pescado en general, sino en el alto consumo de ácidos grasos especiales, en los cuales son particularmente ricos los peces de mares fríos y profundos, como el mar noruego, y en los cuales son singularmente pobres los peces de los ríos y los mares tropicales.

Tamaño de la muestra

Otro aspecto relacionado con la población tiene que ver con determinar el tamaño de la muestra, **o el número de sujetos necesarios** para que, al hacer el estudio, se pueda garantizar que puede encontrarse una diferencia clínicamente significativa con resultado estadísticamente válido. Otro capítulo de este libro presenta en mayor detalle los principios para calcularla (véase el **capítulo 6** "Cálculo del tamaño de la muestra"). A continuación se describirán algunos aspectos pertinentes al caso de los **ECA.**

Aunque de carácter fundamentalmente matemático, el cálculo del tamaño de la muestra requiere que el investigador establezca ciertos elementos clínicos. En concreto, debe conocerse, o aproximarse con cierta precisión, la frecuencia esperada del desenlace en el grupo control, la diferencia entre dicha frecuencia basal y la esperada en el grupo experimental que (en caso de existir) se considere clínicamente importante, y las probabilidades aceptables de cometer errores de tipo I y de tipo II, ya explicados en el **capítulo 9 "Medidas de** frecuencia de asociación y de impacto"[34,35]. La mayor parte de los tipos de experimento descritos antes (cruzados, de equivalencia, secuencial, con asignación por grupos) requieren procedimientos matemáticos particulares, que el investigador debe conocer al planificar su estudio (hay fórmulas diferentes para cada diseño y para cada modalidad de estudio, las cuales pueden encontrarse en el **capítulo 6** "Cálculo del tamaño de muestra").

EJEMPLO

Un grupo de investigadores quiso determinar si la frecuencia de fracasos al tratar a niños menores de 5 años, hospitalizados por neumonía, con amoxicilina oral (tratamiento experimental) era equivalente a la observada con penicilina cristalina parenteral (tratamiento convencional). Según datos previos, de los centros participantes, la frecuencia de fracasos esperada con penicilina era del 10 %. Los investigadores decidieron considerar la frecuencia de fracasos con el tratamiento oral como equivalente a la del tratamiento parenteral si la diferencia observada entre los 2 era ≤ 5 %, con una probabilidad de error I (en una sola cola) de 0,05, y de error II, de 0,9. Con estos datos, se calculó el tamaño de la muestra, a la que se agregó un número adicional de pacientes para permitir tres análisis intermedios. El número final necesario para tener el poder mínimo propuesto (90 %) para encontrar la mínima diferencia propuesta (5 %), con la posibilidad de error por azar máximo que se aceptó de antemano (5 %) fue de 861 sujetos por grupo.

Es necesario tener presente que el cálculo debe hacerse siempre ***antes*** **de empezar el estudio**, pues puede descubrirse más adelante que se han perdido tiempo (planeación, desarrollo del protocolo, entrenamientos, recolección de datos, seguimiento de sujetos) y recursos (salarios, tiempo de investigadores, pruebas de laboratorio, pagos a terceros, programas y bases de datos, asesorías) en un estudio cuyos resultados no son concluyentes por un número muy pequeño de sujetos, el cual, a su vez, no le ofrecía al estudio un poder suficiente para encontrar las diferencias, con significación clínica y estadística, que se buscaban.

Asignación de los sujetos a los tratamientos

Aleatorización

Luego de seleccionar a los sujetos y de hacer una evaluación de entrada, la cual se describe en detalle más adelante en este capítulo, el investi-

gador debe asignar a los sujetos a los grupos del experimento. Como ya se mencionó, el control de dicho proceso por parte del investigador es la característica fundamental que separa los estudios experimentales de los no experimentales. Esta distribución debe dividir a los sujetos en grupos tan similares entre sí como sea posible. De acuerdo con Meinert[36], el proceso de asignación debe cumplir los requisitos mencionados en la **tabla 15.2**.

El método óptimo para hacer la asignación en un experimento clínico es por medio del azar[37,38]. **Una verdadera asignación aleatoria** significa que cada sujeto tiene una probabilidad conocida, usualmente igual para todos los individuos e independientemente de si se recibe o no cada uno de los tratamientos del estudio.

EJEMPLO

Se llevó a cabo un estudio que comparó una intervención psicosocial (escribir sobre situaciones vividas que hubieran generado estrés) con una intervención de control (escribir sobre una situación emocionalmente neutra) en una enfermedad crónica (artritis reumatoide), en busca de mejoría sintomática. Los sujetos fueron aleatorizados, y el método (tabla de números aleatorios) se utilizó de manera que la probabilidad de ser asignado a cada uno de los grupos era del 50 % (conocida), y era la misma para todos los sujetos (también, del 50 %); debe recalcarse que el hecho de ser igual para todos los individuos no significa que deba ser 50 % y 50 %, sino que si la probabilidad para un sujeto de ser asignado al placebo es del 20 %, y, por consiguiente, del 80 % para ser asignado al tratamiento, esas probabilidades deben ser iguales para todos los individuos del estudio. Y, finalmente, en cada sujeto existía siempre la opción de ser aleatorizado a cualquiera de los dos grupos, tanto si el sujeto anterior había quedado en el grupo experimental como si había sido asignado al grupo control (independiente).

La principal ventaja de la aleatorización es que evita los sesgos potenciales que surgirían si la asignación se hiciera de manera directa por parte del investigador. Se trata de un mecanismo automático de control de sesgos, uno de los que dan a los experimentos clínicos su prestigio como diseños de investigación. Se trata de la característica principal del experimento clínico y la que le da su particularidad.

Una vez se decide utilizar un método de asignación por azar, se obtienen varios beneficios, de los cuales el principal es que al asignar los tratamientos al azar se obtienen grupos de tratamiento que son similares, antes de recibir las intervenciones, respecto a factores conocidos o desconocidos en general y en particular respecto a los factores pronósticos que podrían afectar las variables dependientes[39]. **Y si hay diferencias entre** los grupos (lo cual, en efecto, puede ocurrir), tales diferencias se deberán al azar, y no a sesgos. Esta ventaja del diseño de distribución simétrica se obtiene más fácilmente cuanto mayor sea el tamaño de la muestra, si bien hay métodos para mejorar la probabilidad de lograrla en tamaños de muestra menores, como se verá más adelante.

Existe una gran variedad de métodos para la aleatorización, entre los cuales se cuenta el uso de programas en celulares o en computadores para generar números aleatorios, el uso de tablas de números aleatorios que se encuentran a menudo como apéndices en los libros de investigación o de bioestadística, y, finalmente, el empleo de programas específicamente diseñados para aleatorizar[37,40,41]. El apéndice presenta ejemplos de una asignación aleatoria empleando una tabla de números aleatorios y un programa de aleatorización.

Relativamente a menudo, se acude a otros sistemas para asignar que no cumplen los criterios estrictos de distribución por azar, y, por tanto, no son aceptables en un experimento clínico.

La asignación a los grupos de estudio decidida por el investigador viola el primer requisito de la **tabla 15.2, y puede llevar a sesgos si, de manera** consciente o inconsciente —pero sistemática— se asignan más a menudo o menos a menudo a sujetos con mejor o con peor pronóstico (según la edad, la gravedad o la duración de la enfermedad, la presencia de enfermedades asociadas, etc.) a uno de los grupos del experimento.

Tabla 15.2 Requisitos de un sistema de asignación adecuado

1.	Ni el investigador ni el personal del estudio deben **conocer la asignación antes del inicio del tratamiento**
2.	Las asignaciones futuras no deben ser predecibles **con base en las asignaciones previas**
3.	El orden de asignación debe ser reproducible
4.	Se deben documentar los métodos de generación y **administración de la asignación**
5.	El proceso debe tener propiedades matemáticas **conocidas**

Fuente: Modificado de Meinert[35].

Las asignaciones alternadas, donde el primer sujeto recibe un tratamiento, el segundo individuo recibe el otro tratamiento, y, sucesivamente, se van alternando de forma ordenada y predecible, no cumplen la segunda recomendación. Algo similar puede decirse de la asignación a los grupos según el último dígito del número del registro hospitalario o del día de nacimiento (par a un grupo, impar al otro), que no cumple los criterios de probabilidad igual, conocida e independiente.

El empleo de una moneda cumple a cabalidad todos los criterios, si bien es importante enfatizar que debe registrarse el resultado del proceso para que quede un rastro reproducible (para una auditoría posterior, por ejemplo). No es, sin embargo, muy práctico; y si se hace en presencia de los sujetos del estudio, ello puede no ser entendido (¡imagine el lector la expresión del paciente cuando el investigador arroje una moneda al aire para decidir el tratamiento!).

Estratificación

Es más probable que la asignación aleatoria produzca grupos similares cuando se aleatoriza a un gran número de sujetos. Si el tamaño de muestra es pequeño, aumenta la probabilidad de que aparezcan diferencias entre los grupos en algunas características de base, tanto conocidas como desconocidas. Por ello, la asignación aleatoria puede combinarse con dos estrategias que buscan garantizar una mayor similitud de los grupos: 1) la ***estratificación pronóstica*** y 2) la ***aleatorización por bloques***.

En la estratificación pronóstica, la información sobre las características de base de cada sujeto se utiliza para dividir a los participantes en subgrupos de acuerdo con las variables que se consideren importantes por su asociación a las variables dependientes (hipertensión arterial leve, moderada o grave, según la clasificación europea, o fracturas de tipo I, II o III). La aleatorización posterior en cada uno de los subgrupos (llamados ***estratos***) hace que dicha variable de confusión se distribuya homogéneamente entre los grupos de los tratamientos[37,42]. El mismo mecanismo podría emplearse para controlar otras variables de confusión; sin embargo, no es aconsejable abusar de dicho método, pues con cada variable que se agrega aumenta la complejidad del estudio, por cuanto el número total de estratos es igual al producto del número de subgrupos en cada variable.

EJEMPLO

Un investigador considera que el peso al nacer es una variable pronóstica muy importante para la mortalidad neonatal. En consecuencia, decide dividir a los recién nacidos de un experimento en 4 subgrupos: 1) de menos de 1000 g de peso, 2) de entre 1001 y 1750 g, 3) de 1751-2500 g, y 4) de más de 2500 g. Un neonato elegible de 1800 g se aleatorizaría (mediante el método elegido; por ejemplo, tabla de números aleatorios) dentro del tercer estrato; otro, de 1200 g, dentro del segundo, y así sucesivamente. Si el investigador decidiera estratificar también según la vía del parto (vaginal o por cesárea), el número total de subgrupos sería 8 (peso < 1000 g y nacidos por vía vaginal, y peso < 1000 g y nacidos por cesárea, etc.).

Dadas las importantes implicaciones en cuanto a complejidad, la selección de las variables para conformar los estratos debe ser muy cuidadosa, con preferencia por las que sean significativamente más importantes y más fáciles de medir en cada sujeto a su ingreso. En los ensayos multicéntricos, se suele considerar cada centro participante como un estrato independiente. Por último, el empleo de estratificación implica ciertas modificaciones en el análisis[37,42].

La aleatorización por bloques permite evitar disparidades en el número de individuos asignados a los grupos del estudio, particularmente, cuando el tamaño de muestra es pequeño. Si los sujetos van a ser aleatorizados a dos tratamientos, A y B, se hace la aleatorización por pequeños bloques de cuatro, seis u ocho sujetos, y asegurándose de que en cada bloque de tamaño par (cuatro, seis u ocho individuos) la mitad de los sujetos sea asignada al grupo A, y la otra mitad, al B. Este mecanismo garantiza que el número total de sujetos en los grupos nunca diferirá por una magnitud > b/2, donde ***b*** es el tamaño de los bloques, incluso si el estudio debe terminar antes de completarse el reclutamiento de la totalidad de los sujetos[37]. Sin embargo, si el experimento no está enmascarado y el número de los bloques es fijo, los investigadores conocerán la asignación del último sujeto de cada bloque antes de su aleatorización, lo cual, además de violar el primer principio de la **tabla 15.2**, puede llevar a sesgos. Esto puede evitarse si no se revela el tamaño de los bloques y si este se cambia a lo largo del reclutamiento. Una práctica usual es el uso tamaños de bloques

variables y determinados por azar (lo cual hacen automáticamente los programas de computador de aleatorización).

EJEMPLO

En un estudio de evaluación de respuesta hormonal a una intervención en pacientes en diálisis crónica, se necesita a 38 pacientes para encontrar la diferencia que se considera clínicamente importante. Se hace una aleatorización utilizando un programa que escoge al azar bloques de 4, 4, 6, 4, 8, 8 y 4 sujetos, y que distribuye por igual, dentro de cada bloque, a los sujetos entre la intervención experimental y la de control. De esta manera, al final del proceso se tiene a 19 sujetos en el grupo del medicamento, y a 19, en el grupo placebo.

Como puede apreciarse, la forma de la asignación reviste una enorme importancia. La designación del estudio debería ser siempre, como mínimo, **ECA (el *randomized clinical trial*,** o **RCT** de la literatura anglosajona), y debería siempre identificarse como tal, incluso en el título de la publicación. Como se habrá notado, el experimento clínico debe tener un grupo control (experimento clínico controlado), pero la designación de controlado está incluida dentro del término ***aleatorizado*, por lo que no es necesario hablar de experimento clínico controlado y aleatorizado, y será suficiente ECA.**

Dada su importancia, la recomendación general es que se considere una falla potencial grave la ausencia de aleatorización. Para la mayoría de expertos en apreciación crítica, no vale la pena leer un artículo que compare dos o más intervenciones si la asignación de los sujetos a ellas no fue aleatoria[43,44].

Asignación enmascarada (antes denominada ciega)

Además de la distribución aleatoria, el investigador debe garantizar una ***asignación enmascarada*** **(se prefiere este término, el cual** se considera políticamente más correcto). Esto se logra cuando la decisión sobre incluir a un sujeto en el estudio se toma sin conocer el tratamiento que dicho individuo recibiría en caso de participar (el primer principio de la **tabla 15.2**). **Lo contrario puede introducir sesgos,** al crear grupos con diferencias al comienzo del estudio.

EJEMPLO

Un investigador ha preparado una asignación aleatoria mediante un método adecuado, y a partir de ella ha elaborado una lista con los tratamientos que debe recibir cada uno de los sujetos del estudio. Dado que con dicha lista es posible conocer por anticipado el tratamiento que recibirá cada individuo del estudio, este paso puede llevar a sesgo. El investigador puede decidirse por invitar a todos los sujetos (con buen o mal pronóstico) cuando él sabe que recibirán el tratamiento control, y no invitar a los de más pobre pronóstico si el tratamiento que sabe que van a recibir es el experimental. De esta manera, se crean grupos que son diferentes al comienzo del estudio

Este aspecto ha demostrado influir sobre los resultados de los estudios: los experimentos sin asignación enmascarada tienden a sobreestimar el efecto de la terapia experimental al comparar con lo observado en estudios con adecuado enmascaramiento[45,46]. **Los métodos para garantizar** una asignación enmascarada incluyen la aleatorización mediante llamadas telefónicas a un centro coordinador o el empleo de sobres sellados, opacos y numerados secuencialmente, y los cuales contienen el tratamiento y solo deben abrirse una vez el sujeto haya ingresado al estudio. Ello garantiza que solo se puede conocer el tratamiento al cual ha sido asignado cada paciente después de su ingreso al estudio[47,48]. Dicha asignación enmascarada no debe confundirse con la ***evaluación enmascarada de los desenlaces*, que se describe más adelante.**

La racionalidad del experimento clínico aleatorizado

La estrategia más apropiada para evaluar intervenciones en salud es la conducción de experimentos clínicos controlados. En el estudio se ensambla una cohorte, cuyos participantes son asignados aleatoriamente a recibir una u otra de las intervenciones que se quiere comparar, y posteriormente son seguidos y evaluados hasta determinar si se presentan o no los desenlaces de interés.

La comparación de la frecuencia de los desenlaces en los grupos expuestos a las diferentes intervenciones permite establecer el efecto relativo que cada intervención tiene sobre la probabilidad de presentar el desenlace. Si el

efecto es homogéneo en los diferentes participantes, y si otros factores que puedan influir en la probabilidad del desenlace se distribuyen balanceadamente entre los sujetos comparados, se puede atribuir la diferencia en la frecuencia del desenlace a que los participantes recibieron diferentes intervenciones.

Se pueden concebir formas diferentes de la aleatorización para que los grupos que reciben las diferentes intervenciones resulten balanceados respecto a factores pronósticos que pueden modificar del efecto de los tratamientos. Entonces, ¿por qué insistir en que el mecanismo de asignación sea aleatorio? La respuesta es que solo mediante asignación aleatoria se puede hacer una estimación no sesgada del efecto de las intervenciones. Pero, ¿qué se quiere decir con ***estimación no sesgada***, y cómo se conecta dicho concepto con el proceso de aleatorización? Para entenderlo, hay que considerar cómo funciona el proceso de estimación de parámetros de poblaciones con base en la evaluación de muestras.

El problema de estimación se presenta cuando se quiere saber cómo se distribuyen una o varias características (variables) en un conjunto de elementos definido de forma precisa, conocido como población. ¿Cómo se mide la distribución de las variables en dicha población? Hay dos alternativas: 1) se examinan todos los elementos de la población (es decir, se hace un *censo*), o 2) se examina solo una porción de dichos elementos (es decir, una ***muestra***), y, con base en las mediciones obtenidas en la muestra, se intenta establecer el valor de las variables en toda la población.

La medida que resume la distribución de la variable en la población (por ejemplo, el promedio) es el ***parámetro poblacional*** que se quiere estimar. Cuando se selecciona una muestra y se mide en ella la variable de interés, se obtiene un *estimador* del parámetro poblacional. Para que la muestra permita estimar adecuadamente el parámetro, el proceso de medición (que incluye el proceso de selección de la muestra) debe ser válido y preciso. La validez del proceso depende de que la muestra sea una representación adecuada (válida, no sesgada) de la población de la cual se obtiene. Esta cualidad se conoce como *representatividad*.

Tomando en cuenta lo anterior, una muestra es representativa si permite estimar de forma no sesgada los parámetros de interés. La única manera de asegurar la representatividad de la muestra es que sus elementos sean escogidos al azar. Si a través de un muestreo aleatorio se consigue que cada elemento de la población tenga la misma oportunidad de ser incluido, la distribución de cualquier variable medida en los elementos de la muestra tiende a ser equivalente a la distribución de la variable en la población fuente. Por ejemplo, si en una población el 40 % de los sujetos son varones, la mayoría de las muestras aleatorias de esa población tendrán una frecuencia de hombres cercana al 40 %, y solo en pocos casos la frecuencia será muy diferente del 40 %. El muestreo aleatorio incrementa la probabilidad de que la distribución en la muestra represente la distribución en la población fuente, y evita que, de manera sistemática, la distribución del estimador sobreestime o subestime el parámetro poblacional.

EJEMPLO

Se quiere medir la concentración de un soluto en solución. Si está perfectamente disuelto, la concentración del soluto en cualquier fracción de la solución representa bien la concentración en toda la solución. En contraste, si el soluto se sedimenta, su concentración ya no es uniforme: una muestra de la superficie tendrá una concentración sistemáticamente más baja. Si la solución se agita para que el soluto sedimentado se disperse homogéneamente, se consigue que las muestras tomadas de cualquier parte del recipiente tengan una concentración uniforme, que representa la de toda la solución. Dicho de otro modo, la concentración medida en una muestra bien mezclada es un estimador no sesgado de la concentración de toda la solución. En esta analogía, agitar la solución y homogenizar la concentración equivale a hacer un muestreo aleatorio para obtener estimaciones no sesgadas del parámetro poblacional.

Ahora consideremos la medición del efecto de una intervención; por ejemplo, el de un antibiótico para una enfermedad infecciosa. Supongamos que se tiene un conjunto de 100 pacientes con pielonefritis aguda. ¿Cuál es el efecto aparente de administrar el antibiótico A? La mejor manera para saber cuántos de estos pacientes alcanzarán una curación microbiológica (por ejemplo, cultivo negativo a los 3 días de tratamiento) es administrar el antibiótico a todos los pacientes. Supongamos que con eso hay curación en 80 de ellos (80 %). Ahora, ¿cómo saber si el antibiótico A es mejor que el

B en este episodio de pielonefritis en estos 100 pacientes? Realmente es imposible saberlo: la única forma sería disponer de una máquina del tiempo; transcurridos los 3 días del tratamiento A y evaluados sus efectos, se regresaría a los mismos 100 sujetos al instante previo a la iniciación del tratamiento, se les administra el tratamiento B, y al cabo de 3 días se evalúa el efecto sobre el urocultivo. Supongamos que solo en 40 sujetos se hizo negativo. En ese experimento ficticio se podría establecer que la efectividad de A en los 100 casos estudiados es el doble de la de B (80 % *vs.* 40 %).

Lamentablemente, no se dispone de máquinas del tiempo (al menos, no hasta ahora). ¿De qué otra manera se puede comparar el efecto de A y de B sobre la pielonefritis en los 100 pacientes del caso? En vez de tratarlos a todos con los 2 antibióticos, se puede estimar el efecto administrando A a una muestra probabilística (aleatoria) de, digamos, 50 de los 100 pacientes. La proporción de curación encontrada en dicha muestra (supongamos que es del 75 %) es el mejor estimador (no sesgado) que se puede obtener. ¿Cuál sería el efecto si se les hubiese administrado B a los 100 sujetos? De nuevo, en vez de medirlo con total precisión (en los 100 pacientes), se lo estima en la muestra probabilística (y por ello, representativa) de los 50 sujetos que no recibieron el antibiótico A.

Fue un proceso aleatorio el que los incluyó en el grupo de los no escogidos para recibir A. Si se les administra B a estos otros 50 sujetos, se tiene, simultáneamente, el mejor estimador (no sesgado) del efecto de B. De hecho, un profesional en estadística diría que se han obtenido los elementos para hacer una estimación no sesgada de la diferencia o de la razón entre las proporciones de curación asociadas al hecho de administrar A o B en los 100 sujetos examinados. Los resultados obtenidos en ambas muestras aleatorias simultáneas se pueden generalizar de forma legítima a la población fuente, constituida por las 100 personas participantes.

Gracias a este muestreo aleatorio, se evitó sesgo en la asignación a los tratamientos. Adicionalmente, que los sujetos fueran elegidos probabilísticamente permite calcular la precisión de la estimación; o sea, hacer pruebas estadísticas que cuantifiquen la incertidumbre en los estimadores obtenidos atribuible a la variación de muestreo (aleatoria). Para que tenga sentido cuantificar el posible papel del azar en los resultados (la cuantificación de la incertidumbre), es necesario que la selección de la muestra sea probabilística. Finalmente, la selección aleatoria, al ser intrínsecamente no sesgada, tiende a producir grupos balanceados respecto a todas las otras características (variables medidas o no) de los integrantes de la población estudiada, que tengan el potencial de confundir la evaluación de la asociación estudiada.

La asignación aleatoria es el atributo del diseño experimental que más contribuye a confiar en la validez interna de un estudio específico. El problema que se enfrenta ahora es que los resultados obtenidos solo pueden generalizarse a la población estudiada; o sea, a los 100 sujetos que participaron en la investigación. ¿Cómo ayudan estos resultados a decidir cómo tratar adecuadamente a otros pacientes con pielonefritis aguda? En la medida en la que los 100 pacientes que conforman la cohorte estudiada sean una adecuada representación conceptual del problema de la pielonefritis, los médicos que tratan a otros pacientes pueden extrapolar dichos hallazgos a sus propios casos, tal como se discutirá en el aparte sobre extrapolación de los resultados.

Administración de las intervenciones

El siguiente aspecto en la asignación de un experimento clínico es la administración de las intervenciones, o variables independientes, tanto de la experimental como de la empleada en el grupo control. Estas deben describirse de manera completa, dando suficientes detalles para que el lector pueda reproducirlas con exactitud. Si lo que se está evaluando es un medicamento, la descripción debe incluir la dosis, el horario, la forma de administración y la duración total del tratamiento.

La evaluación de procedimientos quirúrgicos implica definir con claridad las técnicas empleadas; sin embargo, en este último caso surgen problemas adicionales, ya que pueden existir "modificaciones personales" en los procedimientos o variabilidad en la experiencia de los cirujanos. Las primeras deben evitarse solicitando a los cirujanos que se ajusten a lo establecido en el protocolo, mientras que el efecto de la segunda puede controlarse mediante el empleo de un periodo de aprendizaje de la técnica antes de la aleatorización o por la distribución aleatoria de quienes hagan la cirugía[49].

Es necesario describir con la misma claridad el tratamiento del grupo control, el cual, en principio, debe ser la mejor terapia usualmente disponible para la condición que padecen los sujetos del estudio. En caso de no ser así, se estarían violando principios éticos fundamentales, pues adrede se negaría a una parte de los participantes la posibilidad de recibir un tratamiento adecuado. Además, se estaría comparando la intervención experimental contra una que no es la óptima y usual, lo cual crearía un sesgo a favor del nuevo tratamiento y limitaría la generalización de los resultados del experimento.

En muchas condiciones médicas no se cuenta con una terapia efectiva contra la cual se pueda comparar la intervención experimental. En tal caso, el investigador debe decidir entre no hacer nada o emplear un placebo. Se define placebo como una intervención inerte (es decir, sin efecto terapéutico intrínseco) e indistinguible del tratamiento activo[7]. Las razones para usar placebos se describen con mayor detalle en la sección de medición, de este mismo capítulo. Sin embargo, el empleo de placebo no necesariamente significa que el grupo control no recibirá un tratamiento activo, pues a menudo se evalúa la adición de un tratamiento.

EJEMPLO

Varios estudios han evaluado el efecto de agregar corticosteroides al manejo de niños con meningitis bacteriana. En estos casos, el grupo experimental recibe el tratamiento usual con antibióticos más el esteroide, mientras que el tratamiento del grupo control sería el antibiótico usual más placebo del esteroide.

La decisión de utilizar placebo en un experimento debe estar claramente regida por la ausencia de un tratamiento disponible, y debe haber sido aprobada por un comité de investigación y ética. Solo en casos muy especiales se permite el uso de placebo para condiciones donde exista una terapia con efectividad demostrada: cuando se investigue una condición menor y el placebo no le signifique al paciente un riesgo de daño serio o irreversible, o cuando sea absolutamente necesario por razones metodológicas sólidas y científicas evaluar la eficacia o la seguridad de una intervención. En todos los casos, sin embargo, debe tenerse la aprobación de un comité de investigación y ética que sea independiente del investigador.

Como es obvio, los resultados del estudio solo serán aceptables si todas las otras intervenciones en el experimento con comparables, excepto las intervenciones que se evalúan (***variables independientes***). Los resultados pueden ser aceptados si las intervenciones evaluadas son administradas en lo que Feinstein llama "ambientes comparables"[50], y que incluye otras intervenciones diferentes de las intervenciones bajo evaluación, pero que pueden influir sobre las variables dependientes. La presencia de diferencias sistemáticas entre los grupos en otros tratamientos se conoce como ***sesgo de cointervención***[51], y puede producir resultados claramente distintos en el grupo que recibe las medidas terapéuticas adicionales. Esta posibilidad se reduce si en el protocolo se garantiza que los tratamientos serán iguales para los dos grupos (exceptuando las intervenciones estudiadas), y si, adicionalmente, se han descrito de forma cuidadosa las intervenciones adicionales que pueden o no aceptarse en el estudio. De nuevo, el empleo de un placebo (que facilita el doble o el triple enmascaramiento) puede reducir la posibilidad de este sesgo, pues tanto los investigadores como los clínicos tratantes desconocerían el grupo al cual ha sido asignado cada individuo, lo que, a su vez, reduce la posibilidad de que un grupo del estudio reciba de manera preferencial tratamientos adicionales.

Evaluación

Evaluación de base

Una vez el sujeto acepta participar, antes de iniciar la intervención es necesario hacer una evaluación de base[52]. Esta información puede obtenerse mediante entrevistas, cuestionarios, exámenes clínicos o pruebas de laboratorio. Tales datos ayudan al investigador a describir la población de su estudio, al tiempo que le permiten a la comunidad científica comparar los resultados del experimento con los de trabajos similares y establecer su aplicabilidad a otros individuos.

Adicionalmente, la información de base permite evaluar la comparabilidad de los grupos del estudio antes de iniciar la intervención, estratificar por factores pronósticos y evaluar la importancia de ciertos factores de riesgo sobre

los resultados, mediante el seguimiento de los sujetos asignados a un grupo del estudio. Es importante obtener la misma información de base en los pacientes elegibles que no aceptan participar en el estudio, con el fin de establecer si dichos individuos presentan diferencias sistemáticas cuando se comparan con los finalmente incluidos en el experimento.

Evaluación de los resultados: seguimiento, estudio abierto o enmascarado

Una vez los sujetos del estudio reciben las intervenciones, es necesario determinar su impacto sobre la condición de base. El primer paso en este punto es definir con claridad las variables dependientes, las cuales son los desenlaces que el investigador utilizará para medir los resultados. Es importante que estas variables sean ***sensibles*** y relevantes[53]. La variable es sensible si permite detectar diferencias entre los grupos con un número relativamente pequeño de sujetos.

EJEMPLO

Un investigador quiere evaluar los efectos de un nuevo tratamiento en pacientes con artritis reumatoide. Aunque clínicamente de importancia, el empleo de la mortalidad como desenlace principal de dicho experimento plantea la necesidad de estudiar a un gran número de sujetos, pues la baja tasa de letalidad de la enfermedad implica el seguimiento de muchos individuos si se quiere observar suficientes fallecimientos en cada uno de los grupos.

En otro ejemplo, también extremo, muchos experimentos de intervenciones en ancianos (como uso de antihipertensivos) no encuentran resultados positivos, pues utilizan como desenlace la mortalidad; se necesita un número muy grande de sujetos en un estudio para demostrar diferencias significativas, pues la probabilidad de muerte es intrínsecamente alta entre estos sujetos. En tales casos, debería utilizarse otro desenlace, como incidencia de ataques vasculares cerebrales, o infartos miocárdicos, o indicadores de calidad de vida.

Por otro lado, la variable escogida como desenlace es relevante si mide un aspecto importante desde el punto de vista de los objetivos del estudio.

EJEMPLO

Un investigador que evalúa un nuevo antihipertensivo contrastándolo con la terapia usual puede comparar los promedios de la presión arterial media observados en los grupos del experimento. Sin embargo, quizá, es de mayor interés clínico determinar la proporción de sujetos hipertensos que al recibir los tratamientos alcanzaron una presión arterial normal, y todavía más relevante, evaluar las diferencias en desenlaces clínicos como infarto del miocardio, hemorragia cerebral o ataques isquémicos transitorios (AIT).

Tan importante como definir las variables dependientes es establecer quiénes serán los observadores que las medirán y cuáles serán los procedimientos. El principal punto para resolver aquí es el potencial de sesgo en la evaluación de las variables dependientes, pues el conocimiento, por parte del evaluador, del tratamiento que recibe cada sujeto puede hacer que el observador, consciente o inconscientemente, influya en la identificación o en el reporte de los eventos relevantes[45,54].

La probabilidad de este sesgo de medición está directamente relacionada con la subjetividad de las variables empleadas: si el desenlace es la mortalidad total, la probabilidad de sesgo es muy baja, ya que la determinación de la muerte no cambia aún si se conoce el tratamiento recibido por el sujeto. Sin embargo, determinar la causa de la muerte es menos preciso y potencialmente variable, según el conocimiento del evaluador sobre la asignación.

Por otro lado, el investigador debe tener en cuenta la bien documentada tendencia de los sujetos a reportar efectos favorables al recibir un tratamiento, independientemente de la eficacia fisiológica de lo que se recibe. Este fenómeno, conocido como ***efecto placebo***, es muy importante en los experimentos donde la variable dependiente incluye desenlaces subjetivos reportados por el paciente, como la severidad del dolor, la calidad de vida o la frecuencia de efectos secundarios[7,45]. En estos casos, será necesario tener en cuenta no solo las expectativas de los evaluadores, sino, también, las de los pacientes participantes.

La mejor estrategia para controlar estos sesgos es mantener enmascarados (ciegos) a los participantes y a los investigadores respecto a la intervención recibida[55,56]. Esto implica que las

maniobras bajo comparación deben ser tan parecidas entre sí como sea posible, y que, en muchos casos (en especial, cuando se están comparando medicamentos) sea necesario emplear placebos.

Como es obvio, los estudios enmascarados son más difíciles de conducir que los abiertos, donde los participantes, los investigadores y los clínicos tratantes conocen la intervención que cada sujeto recibe. Sin embargo, la necesidad de controlar los sesgos potencialmente introducidos por el observador, por el efecto placebo o por cointervenciones hace que en muchos casos sea necesario conducir estudios enmascarados.

En algunos casos debe recurrirse a la estrategia conocida como ***doble simulación***, la cual permite comparar intervenciones muy diferentes, como un medicamento que debe administrarse tres veces al día y se compara con un medicamento de administración única diaria.

EJEMPLO

Si se quiere mantener enmascarada la intervención en un experimento que compara un antibiótico A de uso parenteral (intravenoso) con uno B, de uso oral, para el control de la infección urinaria, puede recurrirse a la doble simulación: un grupo recibirá el antibiótico A intravenoso y un placebo oral del antibiótico B, en tanto que el otro grupo recibirá el antibiótico B por vía oral y un placebo inyectado del antibiótico A.

Además de definir las variables y quiénes serán los observadores, el investigador debe describir los mecanismos empleados para el seguimiento de los participantes[57], el mejor momento para hacer la evaluación, los instrumentos (cuestionarios, imágenes diagnósticas y pruebas clínicas o de laboratorio) que utilizará para medir las variables dependientes[58], los formularios que se emplearán para recolectar la información[59,60] y los mecanismos que se utilizarán para detectar los efectos secundarios de las intervenciones[61].

Además de considerar la posibilidad de sesgo en la evaluación, el investigador debe establecer la variabilidad de las mediciones cuando haya varios observadores (***variabilidad interobservador***), o cuando un mismo observador la determina en momentos diferentes (***variabilidad intraobservador***)[62]. Otros capítulos del presente libro tratan más profundamente estos y otros aspectos de la medición.

Evaluación de adherencia a las intervenciones

También se requiere determinar la adherencia de los sujetos participantes a los tratamientos recibidos, con el fin de descartar la posibilidad de nuevos sesgos. Si la intervención experimental es superior a la convencional, pero produce efectos secundarios que llevan a algunos sujetos a suspenderla, al no recibir tratamiento, tales individuos no presentarán la respuesta terapéutica esperada, situación que reduce la diferencia respecto a la del grupo control. Esto puede llevar al investigador a concluir, erróneamente, que la nueva terapia no es superior a la tradicional, cuando, en realidad, lo que sucede es que parte de los sujetos del grupo experimental no la recibieron. Lo contrario ocurre si los efectos secundarios se presentan con el tratamiento del grupo control.

Para determinar la adherencia, el investigador puede preguntarles a los sujetos si están consumiendo el medicamento, pedirles a los participantes que lleven un diario, pedir a un familiar que haga la vigilancia, hacer conteo de las tabletas sobrantes durante las visitas de control o medir metabolitos del medicamento en muestras de sangre u orina[63]. Más adelante, en la sección de análisis de los resultados, se discutirán las ventajas y las desventajas de incluir o excluir del estudio a los sujetos que tienen una pobre adherencia al tratamiento.

Análisis

Después de recolectar la información, es necesario proceder a su análisis. En el caso de los ECA, este análisis tiene tres objetivos principales: 1) resumir la información de base de la población del estudio y establecer la similitud entre los grupos del experimento; 2) probar la hipótesis del estudio, al establecer si existen diferencias clínica y estadísticamente significativas entre las variables dependientes, y 3) controlar los posibles efectos de variables de confusión que por azar (si hubo aleatorización) hayan quedado distribuidas diferencialmente entre los grupos de estudio.

Comparación de base

El primer análisis de la información recolectada en el estudio va dirigido a confirmar si el mecanismo de asignación produjo grupos que

fueran similares antes de iniciar las intervenciones; especialmente, en factores que pueden modificar la frecuencia del desenlace. Este análisis se basa en la información recolectada en la evaluación de base descrita previamente. Si se hizo aleatorización, es muy probable que los grupos sean comparables. La comparabilidad debe exigirse en las variables que sean pertinentes al desenlace. Si se encuentra una diferencia de 7 años de edad (promedio de edad: 48 años en un grupo, y 55 en otro) entre los grupos experimental y de control, esta diferencia no será relevante en un estudio sobre una enfermedad dermatológica, pero puede ser muy relevante si se trata de un estudio sobre frecuencia de angina o muerte súbita.

La mayor parte de los autores presenta el resultado de este análisis en la primera tabla de su reporte (generalmente, la **tabla 15.1**). Dichos resultados les permiten al investigador y al lector establecer las características de la población estudiada, lo que es de gran utilidad para determinar la aplicabilidad posterior de los hallazgos, así como determinar si se logró un adecuado control de las variables de confusión al asignar las intervenciones.

Comparación de los resultados: análisis por protocolo o por intención de tratar

En esta fase del análisis, el investigador compara los grupos en términos de las variables dependientes o de interés. Este es el componente del análisis que determina si hay una asociación entre las intervenciones y los desenlaces del experimento. Sin embargo, antes de proceder, el investigador debe establecer la naturaleza de la pregunta planteada por su estudio, la cual puede relacionarse con eficacia (¿puede el medicamento A reducir la presión arterial más que el medicamento B, en condiciones ideales?); o con efectividad, que involucra el cumplimiento de recomendaciones y la adherencia al tratamiento (¿puede el medicamento A reducir más la presión arterial que el medicamento B, en condiciones reales?), y, finalmente, con eficiencia, expresada como el balance entre riesgos y beneficios (¿produce el medicamento A más efectos benéficos que perjudiciales en sujetos hipertensos cuando se lo compara con el medicamento B?)[64-66].

Aunque sutiles, las mencionadas diferencias son muy importantes. El experimento **explicativo** busca establecer la eficacia de la intervención, bien sea para definir su mecanismo de acción o para determinar si esta funciona en condiciones ideales y restringidas. Por el contrario, el experimento **de manejo (también llamado *experimento pragmático*, o *de intención de tratar*) intenta** determinar la efectividad del tratamiento describiendo sus consecuencias (buenas y malas) al emplearlo en una enfermedad y determinando su funcionamiento, según las condiciones clínicas usuales del cuidado de los pacientes.

EJEMPLO

Un experimento explicativo evaluaría la eficacia de una intervención terapéutica (por ejemplo, meditación trascendental) agregada a sujetos hospitalizados, en reposo, que cumplen una dieta estricta y reciben el medicamento con un horario establecido y bajo supervisión. En contraste, un experimento de manejo evaluaría la efectividad de la adición de un medicamento adicional a sujetos hipertensos con tratamiento ambulatorio, que no siempre cumplen las recomendaciones terapéuticas y, en algunos casos, no toman el medicamento por omisión o por sus efectos secundarios.

Desde el punto de vista práctico, estos experimentos pueden diferir en cuanto al tipo de sujetos y de eventos finales que se incluirían en el análisis. El estudio explicativo analiza tan solo a los individuos que recibieron el manejo asignado, y excluye los casos con pobre adherencia, así como los desenlaces observados después de la asignación, pero antes de iniciar el tratamiento, o antes de que el sujeto hubiera recibido la intervención durante el periodo esperado para que esta ejerza su efecto. Tal estrategia analítica también se conoce como ***análisis por protocolo***.

En contraste, el experimento de manejo incluye a todos los participantes (aun a quienes tuvieron pobre adherencia), así como todos los eventos que ocurren después de la aleatorización. Este tipo de análisis se conoce como *análisis por intención de tratar*, **pues lo que se** evalúa aquí no es el tratamiento en sí, sino el resultado de la intención que el investigador tuvo de ofrecer la intervención a los participantes, y la cual tiene en cuenta la adherencia y el cumplimiento por parte del paciente.

Otro de los aspectos que deben considerarse al planear el análisis es la necesidad de llevar a

cabo análisis interinos y de establecer criterios para detener tempranamente el estudio[67,68]. La conducción de un experimento plantea la posibilidad de negar una terapia benéfica a los sujetos participantes, o, alternativamente, su exposición a tratamientos inútiles o, incluso, peligrosos. Por ello, los estudios a largo plazo usualmente tienen de por medio un comité de revisión externo, cuya función es evaluar periódicamente la evidencia que proviene del experimento, para no sobrepasar ciertos límites éticos razonables[67].

No obstante lo anterior, la repetición del análisis estadístico o su aplicación a más de una variable aumenta la posibilidad de cometer un error tipo I, por el cual el investigador podría concluir que hay diferencias entre los tratamientos, cuando estas realmente no existen. Para disminuir la probabilidad de este error, el investigador, por lo común, incrementa el tamaño de su muestra, lo que, es obvio, aumenta las dificultades del experimento. Como en muchos otros casos, la necesidad de dicho análisis intermedio debe evaluarse de forma individual.

Por otro lado, como a menudo varios grupos de investigadores tratan de contestar de manera independiente la misma pregunta, el investigador debe prever la posibilidad de terminar su estudio tempranamente si la publicación de los resultados de otros estudios simultáneos crea problemas éticos[69,70].

Finalmente, en esta sección del protocolo el investigador debe determinar el manejo metodológico de los sujetos admitidos incorrectamente al experimento, así como quienes se pierden durante el seguimiento[71]. La gran preocupación es que la manipulación de dichos sujetos puede crear sesgos.

Las opciones en el primer caso fueron discutidas previamente, al tratar la población del estudio. El manejo de la segunda situación es más complejo y controvertido y, quizá, la decisión final deba tener en cuenta sus causas, que generalmente son: 1) la presencia de efectos secundarios, 2) el empeoramiento en la enfermedad del individuo, 3) el desarrollo de algún otro evento, 4) el empleo de otra intervención, 5) la pérdida de interés por parte del sujeto o de su médico, y 6) el cambio de residencia.

Gent y Sackett[71] sugieren que si un individuo se pierde del seguimiento por alguna de las cuatro primeras causas enunciadas, se le considere no como pérdida al seguimiento sino como una *falla terapéutica*, y su desenlace debe incluirse en el análisis dentro del grupo original. Si la razón para la pérdida al seguimiento es alguna de las dos últimas causas descritas, y el estudio está evaluando la eficacia de la intervención, se debe excluir del estudio al individuo. En cualquier caso, lo mejor es tomar todas las precauciones posibles para evitar la pérdida de sujetos por esas dos últimas causas.

Ajuste por variables de confusión

El tercer objetivo del análisis es controlar por variables de confusión que hayan quedado mal distribuidas en el momento de la asignación, y que por el riesgo de introducción de sesgos pueden dificultar la interpretación de los resultados. Estas variables serían detectadas al comparar los grupos según sus características de base, y podrían controlarse mediante estrategias de análisis estratificado o por análisis de regresión múltiple.

No obstante lo anterior, lo ideal es que no sea necesario recurrir a las citadas medidas, que a menudo son vistas con escepticismo por parte de los lectores de los reportes. De nuevo, la asignación por métodos aleatorios, apoyada por estratificación pronóstica y por el uso de bloques es la mejor medida para reducir el riesgo de grupos no balanceados adecuadamente.

Interpretación

Interpretar los resultados de una investigación en salud conlleva obtener las conclusiones que los hallazgos del estudio tienen para los participantes. En el caso de preguntas sobre intervenciones, en las que el experimento aleatorizado es el diseño de elección, ello implica establecer la dirección, la fortaleza y la precisión de la asociación entre las exposiciones y los desenlaces, verificar que la exposición precedió a los desenlaces y evaluar el control de las variables de confusión y de sesgos que sean pertinentes a la pregunta de interés.

El **capítulo 9** "Métodos de frecuencia, de asociación y de impacto"de este libro presenta las estrategias para determinar si hay una asociación entre las intervenciones y los desenlaces, así como su dirección y su precisión. Cuando el desenlace es una variable dicótoma, como ocurre con la mayor parte de las variables dependientes de interés en investigación clínica,

esto se logra obteniendo medidas de asociación como el riesgo relativo o la reducción absoluta del riesgo, con sus correspondientes intervalos de confianza (IC)[72,73].

Cuando la variable dependiente es continua o es el tiempo que transcurre hasta la aparición de un evento, se requieren otras estrategias de análisis. Adicionalmente, se pueden emplear métodos analíticos para pruebas de hipótesis, que también se describen en los capítulos mencionados al comienzo de este párrafo.

El ECA, en contraste con los diseños observacionales, hace mucho más sencillo interpretar los hallazgos para establecer si la exposición precedió a los resultados y si se logró un adecuado control de sesgos; en especial, por variables de confusión. Como se parte de sujetos que aún no han presentado el desenlace de interés y a quienes se asigna a las exposiciones del estudio, este diseño lleva implícito que la exposición precede al desenlace.

Por otra parte, la asignación aleatoria es el instrumento más poderoso para lograr el control de variables de confusión, tanto conocidas como desconocidas, que pueden interferir con la interpretación de los hallazgos. Si el investigador demuestra que los grupos del estudio fueron muy similares en sus características de base (particularmente, en las que se reconocen como variables de confusión), es seguro asumir que estas, así como otras que no se midieron, no son una explicación para los hallazgos del estudio.

Finalmente, las secciones previas del presente capítulo describen las estrategias para el control de otras fuentes de error sistemático al administrar las intervenciones o al medir los resultados.

Extrapolación

La extrapolación de los resultados de un estudio hace referencia a la aplicabilidad de sus hallazgos a sujetos que no participaron en él. En breve, la combinación de factores biológicos, socioeconómicos y epidemiológicos determina si los resultados observados en los participantes de un experimento son aplicables a otros sujetos[74].

Los factores biológicos se refieren a la presencia de diferencias fisiopatológicas en la enfermedad entre los participantes en el estudio y sujetos "con la misma enfermedad", pero que no participaron en él, y que pueden modificar la respuesta al tratamiento (por ejemplo, algunos agentes infecciosos muestran diferencias en cuanto a su sensibilidad a los fármacos), así como a diferencias en el metabolismo de los fármacos, la respuesta inmune o condiciones ambientales que afectan la efectividad y la seguridad de las intervenciones.

Los aspectos sociales y económicos incluyen variaciones en la adherencia a las intervenciones que pueden llevar a modificaciones en la forma como se prescriben o se usan los tratamientos. Tal variabilidad a menudo surge como consecuencia de limitaciones económicas, en el acceso a recursos técnicos, en las habilidades del personal de salud y en idiosincrasias del comportamiento (por ejemplo, en América Latina es común la creencia de que los antibióticos causan anemia, y eso lleva a que, a menudo, los pacientes suspendan su uso antes de lo indicado).

Finalmente, la presencia de otras condiciones mórbidas en algunos sujetos puede afectar la efectividad de la intervención al hacer más difícil el diagnóstico antes de iniciar la terapia, o debido a un aumento en el riesgo de desenlaces adversos. Por ejemplo, la desnutrición en los niños hace más difícil identificar los casos de neumonía y aumenta el riesgo de una pobre respuesta terapéutica.

A partir de lo dicho en los párrafos precedentes, se hace evidente que establecer si los resultados de un experimento son aplicables a otros sujetos es materia de juicio clínico. Como regla general, se sugiere que el lector considere si el sujeto a quien piensa extender los resultados habría podido ser elegido, en caso de haber estado en el centro donde se adelantó la investigación[74]. Si la respuesta es afirmativa, es razonable concluir que los resultados sean aplicables a él. En caso contrario, el lector debe pensar que existen razones de peso por las cuales los resultados no deberían aplicarse a dicho individuo.

Las mencionadas estrategias, sin embargo, no resuelven todas las dificultades. La extrapolación de los resultados de un experimento conlleva considerar dos aspectos relacionados, pero diferentes[75]. El primero es establecer si los resultados del experimento, que corresponden al promedio del efecto de la intervención observado al agrupar a todos los sujetos del estudio, son aplicables a otro individuo en particular. Para ello puede ser útil comparar las características de la población del estudio con las del sujeto en cuestión, como se describió líneas arriba.

En segundo lugar, el clínico y el paciente están de verdad interesados en saber si la intervención funcionará en su caso individual, y para ello la respuesta promedio, obtenida a través del experimento, puede no ser adecuada. En otras palabras, el paciente individual puede no quedar satisfecho si se le informa que en un estudio que comparó 2 intervenciones se observaron resultados positivos en el 80 % de quienes recibieron la intervención experimental y en el 60 % de quienes recibieron la intervención de control. Lo que los pacientes querrían saber es si ellos estarían o no dentro del 80 % o del 60 % que respondió favorablemente. Ello implica definir cuál es el riesgo basal del paciente de acuerdo con sus características demográficas, sociales y clínicas, y compararlo con la información que sobre ello se pueda encontrar en el reporte del experimento. Infortunadamente, en la mayor parte de los casos no es posible alcanzar un grado tan refinado de predicción.

ASPECTOS ÉTICOS

El **capítulo 36** "Principios éticos en investigación clínica y epidemiológica" de este libro describe en detalle los principios éticos de la investigación en salud. Una de sus secciones presenta los argumentos que más debate generan en torno a los aspectos éticos de los ECA. Los párrafos que siguen resumen dichos elementos.

Aún existe controversia respecto a si es ético adelantar experimentos clínicos. Un punto de vista sostiene que el experimento lleva al clínico a asumir un papel de científico que crea conflicto con las obligaciones hacia el paciente, pues al hacer el estudio no es el clínico, sino otro mecanismo (usualmente, el azar), lo que decide el tratamiento que dicho individuo recibirá. Esto es considerado antiético, así los resultados finales beneficien a la sociedad[76]. Tal punto de vista desconoce, sin embargo, que la justificación para el estudio radica en que no se sabe cuál de las dos intervenciones es mejor.

Por otra parte, el punto de vista contrario arguye que el ECA es el instrumento óptimo para evaluar adecuadamente nuevas intervenciones, y que la historia muestra cómo algunas intervenciones utilizadas y aceptadas con base en estudios no experimentales fueron abandonadas después, por inútiles, o, incluso, por peligrosas. Según esto, el experimento clínico protege a los médicos y a los pacientes de intervenciones inefectivas o inseguras. Además, el empleo del azar no plantea problemas éticos mientras no se haya confirmado que uno de los tratamientos en estudio es superior al otro. Tal concepto se conoce como el ***principio de equilibrio clínico***[77].

Otro de los puntos de controversia es el tratamiento que debe recibir el grupo control. En principio, se considera que este debe ser la mejor intervención disponible para el manejo de la condición. Sin embargo, hay dos elementos de discusión. Por un lado, es posible que, debido a razones económicas o técnicas, en muchos lugares no se cuente con el mejor tratamiento disponible. Esto es muy común en las áreas menos desarrolladas. ¿Podrían los investigadores de dichos lugares emplear en el grupo control el tratamiento usual, así sea inferior al ideal, que existe, pero no está disponible? En esos casos, se propone que se utilice como criterio para seleccionar la intervención de control el criterio triple de la mejor intervención disponible y aplicable, lo cual resuelve el problema de no disponibilidad o de imposibilidad técnica o logística de uso, pero insiste en que sea la mejor opción local.

Por otro lado, algunas posiciones recalcan que, desde el punto de vista metodológico, puede ser necesario emplear placebo en el grupo control (incluso, en algunos casos donde se cuenta con una terapia efectiva), siempre y cuando ello signifique riesgos e incomodidades transitorias y de poca importancia para los participantes (por ejemplo, dolor leve, prurito, etc.). La dificultad está en decidir cuándo los riesgos y las incomodidades transitorias son transitorias y de poca importancia para hacerlas aceptables, razón por la cual la decisión debe quedar en manos de un comité de investigación y ética calificado, y en el cual ***no participe*** el investigador.

Todavía hay un elemento más, relacionado con la disponibilidad de la intervención experimental, tras la terminación del estudio, para los sujetos que participaron en el estudio, así como para individuos similares de la misma región. Si se asume que el experimento confirma que la nueva intervención es mejor que la de control, ¿es ético conducir el estudio con un medicamento o una vacuna en una población que, debido a los costos de la intervención, tiene poca (o ninguna) posibilidad de tener acceso a ella luego de terminado el estudio? Incluso si se garantiza la disponibilidad para los participantes,

¿no debería hacerse lo mismo para la población a la cual pertenecen esos sujetos, así no hayan participado en el estudio? No hacerlo podría considerarse una forma inaceptable de explotar a una población vulnerable.

Finalmente, al igual que como ocurre con cualquier otro estudio, al planear un experimento se deben considerar los principios de respeto por las personas, de balance entre riesgos y beneficios y de justicia[78].

El respeto por la autonomía se traduce en la obligación del investigador de obtener consentimiento informado de parte de los sujetos incluidos en el estudio, que debe obtenerse en todos los casos y en un proceso que culmina en una declaración escrita. Para ello, el investigador debe describir los propósitos y los procedimientos del estudio, los riesgos y las incomodidades y el proceso para asignar el tratamiento. Adicionalmente, se le debe garantizar al sujeto que toda la información tiene carácter confidencial y hay completa libertad para retirarse del experimento en cualquier momento, sin ningún tipo de penalidad. Si se prevé alguna remuneración para el participante, se debe demostrar que esta no influye sobre la aceptación. Se recomienda que quien obtenga el consentimiento sea un profesional de la salud no involucrado en el cuidado del paciente, pues la propuesta por parte del médico tratante puede coartar la libertad del sujeto para negarse a participar[79].

Adicionalmente, el investigador debe identificar los riesgos a los cuales podrían exponerse los participantes, así como los beneficios que pueden obtener, y demostrarse que los segundos sobrepasan los primeros.

CONCLUSIÓN

El mejor instrumento disponible para establecer si una nueva intervención en salud produce más beneficios que perjuicios en los sujetos que la reciben es el ECA. Este diseño le permite al investigador observar los resultados de la intervención en un grupo de individuos y comparar su respuesta con la obtenida en un grupo de control que, por lo general, recibe el tratamiento convencional.

El experimento clínico se basa en un principio simple y lógico: la comparación de dos grupos de sujetos idénticos en todo, excepto en el factor en evaluación.

La piedra angular del experimento clínico es la aleatorización, que permite obtener comparabilidad y similitud entre los grupos.

La conducción de un experimento es un ejercicio complejo, pues el estudio requiere planificación y ejecución cuidadosas, que garanticen no solo la validez y la generalizabilidad de sus resultados, sino que los sujetos participantes no correrán riesgos innecesarios.

APÉNDICE. EJEMPLO DE UNA ALEATORIZACIÓN

Este apéndice describe varios métodos para generar una secuencia de asignación aleatoria. Para ilustrar su empleo, se asume que se quiere asignar sujetos a dos intervenciones (A o B), que cada sujeto tendrá la misma probabilidad de recibir cualquiera de ellas y que la asignación de cada uno es independiente de las asignaciones previas.

ASIGNACIÓN SIMPLE MEDIANTE UNA TABLA DE NÚMEROS ALEATORIOS

La mayor parte de los textos de estadística incluyen tablas de números aleatorios en sus apéndices. Alternativamente, los programas para dispositivos móviles (celulares inteligentes, tabletas electrónicas) y varios programas para computador pueden generar secuencias de números aleatorios. Uno de estos es OpenEpi, un conjunto de programas de acceso gratuito producido la Escuela de Salud Pública Rollins de la Universidad de Emory, en Atlanta, Georgia, Estados Unidos (http://openepi.com/Menu/OE_Menu.htm; última consulta: 10 de noviembre de 2014). La función "Random numbers" permite producir una lista de números aleatorios. La lista que se presenta a continuación corresponde a una serie de 360 dígitos aleatorios entre 0 y 9, agrupados en 60 integrales de seis cifras, y fue producida empleando OpenEpi.

190722	171581	115870	196765	193768	127437
151432	111328	185287	133317	107780	153691
171777	174490	154414	106629	108372	164615
181938	146338	155123	136206	180725	171921
130854	101884	115467	116383	149601	149383
171464	157400	171611	164227	117445	169337
180450	162331	162105	130459	177609	134490
119312	161202	139974	151911	169893	139248
182319	109863	126809	168377	145041	123431
146584	127873	112639	129688	155330	123568

Para asignar dos intervenciones, se decide ***a priori*** y de manera arbitraria que los dígitos impares indiquen una de ellas (A, en este ejemplo), y los pares, la otra (B). A continuación se escogen de forma arbitraria en la tabla un lugar de comienzo y la dirección en la que se leerán los números. Comenzando con el primer dígito de la primera fila y la segunda columna, y desplazándose en sentido horizontal, los primeros 12 dígitos son 171581 115870, los cuales indicarían que la secuencia de las intervenciones para los primeros 12 sujetos sería AAAABA AAABAB.

Alternativamente, se podría haber tomado cada integral de 6 dígitos como un número aleatorio, asignando las intervenciones según si su último dígito es par o impar (171581 = A, 115870 = B, y así sucesivamente). Pueden existir incontables estrategias, y en realidad no es muy importante cuál de ellas se emplee. Lo cardinal es que se tome la decisión ***a priori***, que se la respete estrictamente durante el proceso y que quede un registro claro del procedimiento para auditorías posteriores.

Si se quisiera asignar tres intervenciones, se podría decidir que los dígitos 1-3 correspondan a la letra A; los dígitos 4-6, a B, y los dígitos 7-9, a C; en este caso, se omitiría el cero. Usando un procedimiento parecido, se podría generar una secuencia aleatoria, donde la probabilidad de recibir una intervención sea el doble que la de recibir la otra: los dígitos entre 1 y 6 corresponderían a A, mientras que del 7-9 serían B.

ASIGNACIÓN POR BLOQUES USANDO UNA TABLA DE NÚMEROS ALEATORIOS

Como pudo verse en el ejemplo del párrafo anterior, la asignación simple de los primeros doce sujetos produjo dos grupos con tamaño muy diferente (nueve y tres sujetos en las intervenciones A y B, respectivamente). La asignación por bloques puede evitar ese inconveniente. Un bloque de cuatro significa que dos de los sujetos reciben una intervención, y los dos restantes, la otra. Los bloques pueden hacerse de cualquier número par, de forma que la mitad de los individuos queden asignados a una intervención, y la otra mitad, a la otra.

Si se consideran bloques de cuatro individuos, solo hay seis secuencias posibles para asignar las intervenciones:

1: AABB
2: ABAB
3: ABBA
4: BBAA
5: BABA
6: BAAB

A continuación se podrían asignar estas secuencias empleando el mismo método, ya descrito, para la asignación simple, pero usando solo los dígitos entre 1 y 6 y descartando los 6. La secuencia del ejemplo era 171581 115870 196765 193768, etc. Su empleo para construir la secuencia de asignación por bloques llevaría al siguiente resultado: 1 = AABB, y descartar el 7, 1 = AABB, 5 = BABA, y descartar el 8, 1 = AABB, y así sucesivamente. De esta manera, el número de sujetos asignados a las intervenciones nunca diferirá por más de la mitad del tamaño del bloque. Es posible variar al azar el tamaño de los bloques empleando mezclas de bloques de dos, cuatro o seis sujetos.

Los dos métodos descritos antes se pueden emplear para hacer una asignación aleatoria y estratificada. Suponiendo que se quiere controlar la distribución por sexo, se produciría una asignación para las mujeres, y otra independiente, para los hombres.

OTROS PROGRAMAS PARA ALEATORIZAR

Martin Bland, profesor de la Escuela de Medicina del Hospital de San Jorge, en el Reino Unido, ofrece en Internet un ***Directorio de Programas y Servicios de Aleatorización*** (www-users.york.ac.uk/~mb55/guide/randsery.htm; última consulta: 30 de abril de 2015). Dicho directorio describe varios recursos disponibles para producir asignaciones aleatorias, incluyendo programas de distribución libre o comercial, así como organizaciones que ofrecen el recurso de aleatorización y de manejo de datos en **ECA.**

Referencias

1. Rothman KJ, Greenland S, Lash TL. Types of epidemiologic studies. En: Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd. ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2008. p. 87-99.
2. Kramer M. Part I. Epidemiological research design. En: Kramer M. Clinical epidemiology and biostatistics. Berlín: Springer Berlag; 1988.
3. Fletcher RH, Fletcher SW, Fletcher GS. Treatment. En: Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology. The essentials. 5th. ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2014. p. 132-51.
4. Gordis L. Assessing the efficacy of preventive and therapeutic measures: randomized trials. En: Gordis L. Epidemiology. 4th. ed. Philadelphia: Saunders Elsevier; 2009. p. 131-46.
5. Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007;334:349-51.
6. Meada M, Harria HW, Finland M. Treatment of pneumococcal pneumonia with penicillin. New Engl J Med. 1945;232:747-55.
7. Porta M. A dictionary of epidemiology. 5th. ed. Nueva York: Oxford University Press; 2014.
8. Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. BMJ. 2003;326:1083-4.
9. Sibbald B, Roberts C. Crossover trials. BMJ. 1998;316:1719-20.
10. Louis TA, Lavori PW, Baylar JC, et al. Crossover and self-controlled designs in clinical research. New Eng J Med. 1984;310:24-31.
11. Brown BW. The Crossover experiment for clinical trials. Biometrics. 1980;36:69-80.
12. Jones B, Jarvis P, Lewis JA, et al. Trials to assess equivalence: the importance of rigorous methods. BMJ. 1996;313:36-9.
13. Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. Am J Health Syst Pharm. 2010;67:1337-43.
14. Healy MJR. Distinguishing between "no evidence of effect" and "evidence of no effect" in randomised controlled trials and other comparisons. Arch Dis Child. 1999;80:210-3.
15. Detsky AS, Sackett DL. When was a negative clinical trial big enough? How many patients you needed depends on what you found. Arch Intern Med. 1985;145:709-12.
16. Mayosi BM, Ntsekhe M, Bosch J, et al. Prednisolone and mycobacterium indicus pranii in tuberculous pericarditis. New Engl J Med. 2014;371:1121-30.
17. Friedman LM, Furberg CD, DeMets DL. Basic study design. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials 3rd ed. New York: Springer; 1998. p. 41-60.
18. Sedgwick P. Cluster randomised controlled trials. BMJ. 2012;345;e4654.
19. Correa JC, Pinto D, Salas LA, et al. A cluster-randomized controlled trial of handrubs for prevention of infectious diseases among children in Colombia. Rev Panam Salud Publica. 2012;31:476-84.
20. Armitage P. Sequencial methods. En: Armitage P. Statistical methods in clinical research. Oxford: Blackwell Scientific Publications; 1974. p. 415-25.
21. MCR Vitamin Study Research Group. Prevention of neural tube defects. Lancet. 1991;338: 131-7.
22. Kopec JA, Abrahamowicz M, Esdaile JM. Randomized discontinuation trials: utility and efficiency. J Clin Epidemiol. 1993;46:959-71.
23. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. New Engl J Med. 1991:324:781-8.
24. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med. 1984;3:409-22.
25. The Magpie Trial Collaborative Group. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: a Randomized Placebo-Controlled Trial. Lancet. 2002;359:1877-90.
26. Heart Protection Study Collaborative Group. MRC/BHF Heart protection study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controled trial. Lancet. 2002;360:7-22.
27. Guyatt G. Determining optimal therapy-randomized trials in individual patients. New Engl J Med. 1986;314:889-92.
28. Sedgwick P. What is an "n-of-1" trial? BMJ. 2014;348:g2674.
29. Friedman LM, Furberg CD, DeMets DL. What is the question? En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials,3rd ed. New York: Springer; 1998. p. 16-29.
30. Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ. 2010;340:c117.
31. Friedman LM, Furberg CD, DeMets DL. Study population. En: Friedman LM, Furberg CD, DeMets

DL. Fundamentals of clinical trials 3rd ed. New York: Springer; 1998. p. 30-40.
32. Lipsky PE. Artritis reumatoide. En: Branwald E, Fauci AS, Kasper DL, et al. Principios de medicina interna, 15ta ed. Madrid: McGraw Hill; 2001. p. 2255-65.
33. Friedman LM, Furberg CD, DeMets DL. Issues in data analysis. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials 3rd ed. New York: Springer; 1998. p. 284-322.
34. Friedman LM, Furberg CD, DeMets DL. Sample size. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials 3rd ed. New York: Springer; 1998. p. 94-129.
35. Meinert CL. Sample size and power estimates. En: Meinert CL. Clinical trials. Design, conduct and analysis. Oxford: Oxford University Press; 1986. p. 71-89.
36. Meinert CL. Essential design features of clinical trials. En: Meinert CL. Clinical trials. Design, conduct and analysis. Oxford: Oxford University Press; 1986. p. 67-8.
37. Friedman LM, Furberg CD, DeMets DL. The randomization process. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials, 3rd ed. New York: Springer; 1998. p. 61-81.
38. Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? BMJ. 1999;318:1209.
39. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. J Clin Epidemiol. 1999;52:487-97.
40. Altman DG, Bland JM. How to randomize. BMJ. 1999;319:703-4.
41. Roberts C, Torgerson DJ. Randomisation methods in controlled trials. BMJ. 1999;317:1301.
42. Kernan WN, Viscoli CM, Makuch RW, et al. Stratified randomization for clinical trials. J Clin Epidemiol. 1999;52:19-26.
43. Guyatt G, Rennie D, Meade M, et al (editors). Users´ guides to the medical literature: essentials of evidence-based clinical practice. 2nd ed. Nueva York: McGraw-Hill; 2008.
44. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine. How to practice and teach EBM, 4th ed. Edinburg: Elsevier Churchill Livingstone; 2011.
45. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273:408-12.
46. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609-13.
47. Altman DG, Schulz KF. Concealing treatment allocation in randomized trials. BMJ. 2001;323:446-7.
48. Sedgwick P. Allocation concealment. BMJ. 2012;344:e156.
49. McCulloch P, Taylor I, Sasako M, et al. Randomised trials in surgery: problems and possible solutions. BMJ. 2002;324:1448-51.
50. Feinstein AR. Clinical biostatistics. III: The architecture of clinical research. Clin Pharmacol Therap. 1970;11:432-41.
51. Sacket DL. Bias in analytic research. J Chron Dis. 1979;32:51-63.
52. Friedman LM, Furberg CD, DeMets DL. Baseline assessment. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials, 3rd ed. New York: Springer; 1998. p. 130-9.
53. Roland M, Torgerson D. What outcomes should be measured? BMJ. 1998;317:1075.
54. Gotzche PC. Bias in double blind trials. Danish Med Bull. 1990;37:329-36.
55. Friedman LM, Furberg CD, DeMets DL. Blindness. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials 3rd ed. New York: Springer;1998. p. 82-93.
56. Day SJ, Altman DG. Blinding in clinical trials and other studies. BMJ. 2000;321:504.
57. Greenberg G. Logistics and management. En: Chaput de Saintonge DM, Vere DW. Current problems in clinical trials. Oxford: Blackwell Scientific Publications; 1984. p. 29-34.
58. Meinert CL. Data collection considerations. En: Meinert CL. Clinical trials. Design, conduct and analysis. Oxford: Oxford University Press; 1986. p. 119-37.
59. Wright P, Haybittle J. Design of forms for clinical trials. BMJ. 1979;2:529-30.
60. Kmatternud GL, Forman SA, Canner PL. Designs of data forms. Controlled Clin Trials. 1983;4:429-49.
61. Friedman LM, Furberg CD, DeMets DL. Assessment and reporting of adverse events. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials, 3rd ed. New York: Springer; 1998. p. 170-84.
62. Friedman LM, Furberg CD, DeMets DL. Data collection and quality control. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials, 3rd ed. New York: Springer; 1998. p. 156-69.
63. Boudes P. Drug compliance in therapeutic trials: a review. Control Clin Trials. 1998;19:257-68.

64. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. J Chron Dis. 1967;637-48.
65. Sacket DL. The competing objectives of randomized clinical trials. New Engl J Med. 1980;303:1059-60.
66. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials. 2009;10:37.
67. Friedman LM, Furberg CD, DeMets DL. Monitoring response variables. En: Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials, 3rd ed. New York: Springer; 1998. p. 246-83.
68. Séller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. Biometrics. 1987;43:213-23.
69. Laupacis A, et al. How should results from completed studies influence ongoing clinical trials? Ann Intern Med. 1991;115:818-22.
70. Fletcher A, Spiegelhalter D, Staessen J, et al. Implications for trials in progress of publication of positive results. Lancet. 1993;342:653-7.
71. Gent M, Sacket DL. The qualification and disqualification of patients and events in long term cardiovascular clinical trials. Thromb Haemot. 1979;41:123-34.
72. Lapaucis A, Sackett DL, Roberts RS. An Assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318:1728-33.
73. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analysis of randimozed trials. J Clin Epidemiol. 1994;47:881-9.
74. Dans AL, Dans LF, Guyatt GH, et al. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. JAMA. 1998;279:545-9.
75. Mant D. Can randomised trials inform clinical decisions about individual patients? Lancet. 1999;353:743-6.
76. Hellman S, Hellman DS. Of mice but not men. New Eng J Med. 1991;324:1585-9.
77. Freedman B. Equipoise and the ethics of clinical research. New Eng J Med. 1987;317:141-5.
78. Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects. Geneva: CIOMS; 2002.
79. Colombia, Ministerio de Salud. Resolución 8430 de 1993 Por la cual se dictan normas científicas, técnicas y administrativas para la investigación en salud. Bogotá: Instituto Nacional de Salud; 1993.

CAPÍTULO

16 Estudio de casos y controles

Carlos Gómez-Restrepo

INTRODUCCIÓN

Los estudios de casos y controles (CC) son un tipo de estudio observacional analítico que surge durante el siglo XIX, como medio para estudiar factores de riesgo de las enfermedades prevalentes para ese momento, y más adelante, para evaluar factores de protección.

El estudio, clásicamente, inicia con la identificación de personas con enfermedad u otro desenlace de interés y un buen grupo de control sin la enfermedad o el desenlace de estudio con el cual se compara. El estudio tradicional de CC, basado en casos, es el estudio donde el muestreo de los casos surge *al final* de la potencial cadena de causalidad, por lo que implica precauciones con las inferencias causales que se hagan.

Tradicionalmente, en el estudio de CC los casos se seleccionan de acuerdo con la presencia del desenlace o la enfermedad de interés y los controles sin esa presencia, para posteriormente investigar sobre la exposición a los factores de riesgo o de protección que se propongan. Cabe anotar que el CC tiene la ventaja de investigar múltiples factores a la vez, lo cual es de gran ayuda al momento de investigar la multicausalidad.

Este diseño es adecuado para investigar trastornos raros, con periodos largos de latencia; demanda muestras menores que las requeridas en otros estudios (cohortes), puede estimar el riesgo relativo (RR) a través del OR —por las siglas en inglés de *odds ratio*— (RRI) e investiga múltiples exposiciones, medidas de forma transversal o retrospectiva. A estas ventajas deben agregarse sus limitaciones principales, pues son susceptibles de sesgos de diversa índole y puede haber problemas con las inferencias temporales a las que se lleguen.

HISTORIA

El origen de los estudios de CC se remonta al siglo XIX; en particular, como método para proveer respuestas rápidas a las enfermedades prevalentes en dicho momento y a las epidemias que se presentaban a finales de ese siglo y a comienzos del siglo XX[1]. El desarrollo de los métodos en sociología durante la primera mitad del siglo XX apoyó el inicio de estudios acerca de las causas de las entidades nosológicas partiendo de los desenlaces (vale decir, de las enfermedades), y buscando retrospectivamente los factores desencadenantes (riesgo-exposición) o protectores de estas[2].

Durante el siglo XX, se avanzó muchísimo en la epidemiología de las enfermedades, y empezaron a

cobrar cada vez mayor importancia las enfermedades crónicas, así como a disminuir su importancia las agudas de carácter infeccioso. Ello hizo que tomara mayor relevancia el estudio de enfermedades con tiempos de latencia prolongados[3], con mayor número de factores de riesgo asociados y con mayores costos asociados a la investigación. Dichos factores llevaron al surgimiento y el desarrollo del diseño de CC, que, por sus características particulares, se adaptaba con facilidad a problemas que requerían una respuesta expedita, confiable y económica, pero que, también por sus características, no podían ser resueltos de otra forma.

A comienzos del siglo XX, empezaron a aparecer estudios, ya clásicos, que muestran el desarrollo de este tipo de diseño. En 1920, Broders[4] tomó a un grupo de 537 pacientes con epitelioma de células escamosas del labio (casos) y uno de 500 controles sin epitelioma (controles), e indagó retrospectivamente acerca del hábito de fumar, tan común por ese entonces. A dicho estudio le siguieron otros que utilizaron la misma metodología, como el realizado en 1926 por Janet Lane-Clayton, quien comparó a 500 mujeres hospitalizadas con cáncer de seno y a 500 mujeres controles sin cáncer de seno. A cada una de las participantes de esos grupos, las interrogó sobre el número de hijos que tenían; las primeras reportaron 3,5 hijos y las segundas 5,5 hijos, número bastante alto para los estándares actuales. En el mismo estudio, Lane-Clayton, al interrogar otros factores, sugirió el *sesgo de recuerdo* en el que podrían haber incurrido las mujeres con cáncer de seno, y su posible influencia sobre la validez del estudio[5].

Más adelante, en 1928, Lombard y Doering reportaron un estudio con individuos diagnosticados con cáncer (casos) y no diagnosticados con dicha enfermedad (controles), e hizo un emparejamiento, o apareamiento (*matching*), por sexo y edad, y utilizando las historias clínicas que tenía[6]. Tal metodología ha tenido amplia aceptación, y aunque su uso en estudios de CC ha sido discutido, se le reconocen algunos usos muy precisos, que se verán más adelante en este capítulo. También el diseño de CC se utilizó para estudiar la exposición intrauterina al dietilestilbestrol en adolescentes con cáncer vaginal. En otro campo, Gregg utilizó la misma metodología para indagar la exposición a la rubeola en pacientes que presentaban cataratas congénitas[7].

Durante la primera mitad del siglo XX, se dio el ingreso de la industria tabacalera a los países occidentales, y rápidamente la mayoría de los hombres se convirtieron en asiduos consumidores de cigarrillo. Para 1920, el argentino Angel Roffo, basándose en estudios de laboratorio con animales, postuló la posible asociación del cigarrillo al cáncer de pulmón[8]. A esos hallazgos le siguieron otros que, finalmente, derivaron en tres estudios clásicos de CC que se publicaron en 1950; dos de ellos, realizados en Estados Unidos: el primero de ellos, por Wynder y Graham[9], y el segundo, por Levin, Goldstein y Gerhardt[10]; hubo uno más en Inglaterra, por Doll y Hill[11]. En todos ellos se estableció, de manera definitiva, la asociación entre el hábito de fumar y el carcinoma de pulmón.

Las mencionadas investigaciones, y otras más, sembraron las bases para el diseño de CC, que aún hoy sigue generando respuestas requeridas en las ciencias de la salud. Por otra parte, se empezó a utilizar ese diseño para brotes epidémicos, estudios de factores de riesgo y protección, así como de etiología de enfermedades, y paulatinamente reemplazó al estudio de cohortes; en especial, a las cohortes concurrentes, para evitar el seguimiento de pacientes y ahorrar costos, lo que brindó mayor rapidez y eficiencia a la investigación[12-15].

Paralelo a este desarrollo, se dieron otros grandes avances en lo metodológico. En 1951, Jerome Cornfield demostró cómo el RR se puede aproximar en los estudios de CC mediante la *razón de suertes* (o momios; en inglés, *odds*) en los sujetos expuestos entre los casos sobre la razón de suertes en los sujetos expuestos entre los controles. Se estableció la medición conocida como razón de suertes (o de momios), o *indirecto*, RRI (*odds ratio* [OR]), una medida de gran utilidad para establecer asociaciones causales[16].

Por otra parte, y tomando en cuenta el teorema de Bayes, se transformó la razón de *odds* de exposición en CC por una razón de probabilidades condicionales de enfermedad en pacientes expuestos y no expuestos; esto es, por una razón de riesgos cuando la probabilidad de la enfermedad en la población era baja (lo que en términos prácticos se ha establecido como una prevalencia de enfermedad o desenlace < 10 %).

Se planteó, entonces, como condición necesaria para aproximar el RRI (OR) al RR, el *supuesto de enfermedad rara* (en inglés, *rare disease assumption*). Cornfield propuso, además, que para que la aproximación fuese válida se requería que los

casos fueran representativos de todos los casos, y, así mismo, los controles representaran a todos los controles en la población blanco.

En 1951, el estadístico Edward H. Simpson describió, en un estudio sobre la interacción de dos variables, la conocida ***paradoja de Simpson***, según la cual el OR crudo es diferente de OR similares de los estratos estudiados, por lo cual cabe concluir que una tercera variable de estratificación, conocida posteriormente como *de confusión*, se podría dejar por fuera de los modelos matemáticos, si estuviese relacionada con el desenlace en los pacientes no expuestos y sin la exposición en los controles (***no desenlace***). Tal hallazgo sirvió de base para describir, en 1957, la variable de confusión, por Kish[17].

Más adelante, este diseño se vio enriquecido con los reportes sobre sesgos, como el de sobrevida (Sesgo de Neyman) y el de mala clasificación, que tienen enorme importancia en los estudios de CC. Como un ejemplo del primer sesgo mencionado, al observar a pacientes con tuberculosis (casos) se encuentra que tienen frecuencias menores de cáncer pulmonar que los controles; sin embargo, es claro que al fallecer tempranamente, los pacientes con tuberculosis (TBC) no tuvieron tiempo para desarrollar cáncer[18].

Como ejemplo del segundo sesgo, el de mala clasificación, un estudio encontró que hasta el 56 % de las personas circuncidadas no lo reportaban (falsos negativos), y que el 18 % fueron falsos positivos[19].

El diseño de CC es una herramienta de primer orden al momento de hacer una investigación, por la simplicidad de su diseño, por la opción de obtener un número significativo de pacientes en el corto plazo y por la posibilidad de estudiar simultáneamente varios factores de riesgo.

GENERALIDADES

El diseño de CC es una alternativa de menor costo y de mayor rapidez que el estudio de cohortes para investigar la asociación exposición-enfermedad, así como a la hora de identificar factores de riesgo para enfermedades o desenlaces raros (bastante usuales en medicina). Tiene gran utilidad para evaluar simultáneamente varios factores de riesgo o de protección entre todos los factores cuya asociación al desenlace de interés se quiera conocer.

El estudio de CC pertenece al grupo de *estudios observacionales analíticos*, llamados así porque describen aspectos en los que el investigador no interviene (si se quieren buscar factores de riesgo para la infección de una herida quirúrgica en el ámbito hospitalario, el investigador no asigna quién tendrá o no infección, y tampoco tiene injerencia en cuanto a los potenciales factores de riesgo, como la edad, la diabetes, las comorbilidades, el tipo de cirugía, etc.). Por esto se lo conoce como observacional, y, dado que se hace una comparación de los hallazgos en el grupo de casos con los hallazgos en el grupo de controles, se lo clasifica como analítico.

El estudio de CC compara la exposición a un factor asociado, o potencialmente causal, de riesgo o de protección entre un grupo que ya tuvo el desenlace (casos) y un grupo que no lo tuvo (controles).

Hennekens y Buring (1987) definen el estudio de CC como un diseño observacional analítico donde los sujetos son seleccionados sobre la base de la presencia de una enfermedad o un desenlace (casos) o la ausencia de estos (controles), y posteriormente compara la exposición de cada uno de dichos grupos a uno o más factores o características de interés[20].

Dicho de otra forma, es un estudio que inicia con la identificación de personas con un desenlace o una enfermedad de interés (casos) y se compara con un grupo de personas sin el desenlace o la enfermedad de interés (controles).

Tomando en cuenta lo referido en la definición, se puede plantear gráficamente este diseño como se ve en la **figura 16.1**.

El estudio se inicia con dos grupos:

- Casos (presentan la enfermedad o el desenlace de interés).
- Controles (no presentan la enfermedad o el desenlace de interés).

En cada uno de estos grupos se investiga la presencia o la ausencia de factores de riesgo o de protección que precedieron a la aparición del desenlace (enfermedad), a través de las pruebas que ya se tengan recolectadas (laboratorios, descripciones en la historia clínica, escalas, etc.), entrevistas o cuestionarios que se puedan hacer, y desarrollados de igual forma para los dos

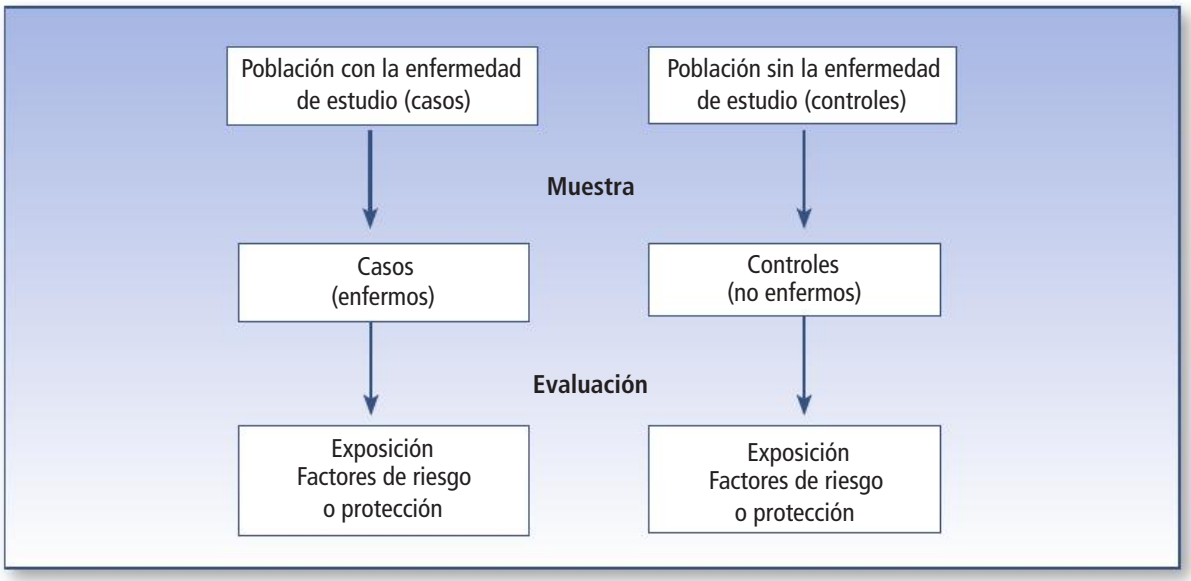


Figura 16.1 Diseño de cc.

grupos (casos y controles). Este tipo de estudio permite, como ya se ha señalado, indagar o evaluar un gran número de factores asociados, que permiten determinar posibles etiologías y posibles factores de riesgo, así como indagar acerca de factores de protección, lo cual constituye una de las principales ventajas de tal diseño.

Algunas personas se refieren al diseño de cc como un estudio retrospectivo, pues en él se investigan los factores de exposición cuando ya se tiene el desenlace. Sin embargo, dicho término es inadecuado y no se recomienda su uso, pues, a pesar de que la recolección de la información sobre la exposición se hace, en general, de forma retrospectiva (en bases de datos, en historias clínicas completas o en otras fuentes de información), también puede hacerse de manera prospectiva, por medio de entrevistas, de cuestionarios o de otros métodos que procuren la "reconstrucción de los hechos".

EJEMPLO

Para investigar los factores de riesgo en el estudio mencionado, de infección de herida quirúrgica, puede acudirse a las dos estrategias: se seleccionan los casos (pacientes infectados) y los controles (no infectados), y retrospectivamente se buscan en sus historias clínicas los datos (edad, sexo, clasificación de riesgo quirúrgico con la escala ASA, presencia de diabetes, tipo de cirugía, sala en la que se hizo el procedimiento, quién fue el cirujano y cuántas personas ayudaron o intervinieron, etc.).

Como dificultades en el tiempo mencionado, puede encontrarse que la clasificación de ASA no se halla disponible en todos los pacientes, que no es claro si algunos pacientes que fueron sometidos a cirugías de urgencia tienen o no tienen diabetes, etc. Pero también se puede solicitar la información de manera estandarizada a todo paciente que se someta a cirugía (sobre las características que se quieren estudiar), y a medida que se les da de alta se los clasifica como *casos* si tuvieron infección, o como *controles* si no la tuvieron. Como se ve, en este caso es más probable que la información esté completa y sea correcta, pues se obtuvo de manera estandarizada y prospectiva.

Vale la pena recordar, pues, que la recolección de los datos acerca de los factores de riesgo o de protección estudiados se puede hacer de manera retrospectiva (bases de datos, historia clínica, etc.) o de manera prospectiva, mediante métodos (entrevistas, cuestionarios, etc.) que procuran reconstruir los factores asociados o causales que estaban o no presentes antes de que la persona presentara (casos) o no (controles) el desenlace de interés. Por tanto, los términos *retrospectivo* o *prospectivo* no hacen referencia al tipo de diseño, sino a la manera de recolectar los datos (factores asociados o causales) que se requiere estudiar.

Como puede imaginarse, en la recolección retrospectiva puede encontrarse el investigador con vacíos de información, con datos incompletos o imposibles de clasificar. Al revisar datos

de historias clínicas, puede ocurrir que no esté registrada la raza sino en el 15 % de los casos, de manera que el dato no podrá ser evaluado.

EJEMPLO

En un estudio se plantea la siguiente pregunta: *¿Cuáles son los factores de riesgo y de protección que se asocian a la presentación de un episodio depresivo en mujeres entre los 35 y los 55 años?* Para dicho estudio, se reúne en una base de datos de su hospital a una serie de pacientes con el mencionado diagnóstico y que cumplan los criterios de inclusión y de exclusión previamente definidos (por ejemplo, que cumplan los criterios diagnósticos de la Clasificación Internacional de las Enfermedades (CIE-10) y del *Diagnostic and Statistical Manual of Mental Disorders*-V (DSM-V) y con las características sociodemográficas previamente definidas (edad, sexo, estado civil, raza, etc.). Así mismo, se definen los controles que podrían ser recuperados aleatoriamente de la misma base de datos con alguna otra enfermedad o sin enfermedad (controles). Definidos ambos grupos, se recuperan en ambos los datos acerca de la presencia o no de los factores de exposición que se haya definido estudiar. Algunos de ellos podrían ser: actividad laboral, pérdida de seres queridos durante el año inmediatamente anterior a la presencia del desenlace (depresión), antecedentes familiares de enfermedad afectiva, presencia de actos violentos hacia algún familiar en primer grado, tipo de dieta, antecedentes quirúrgicos a lo largo de los últimos 12 meses, etc.

TIPOS DE ESTUDIOS DE CC

Los estudios de CC pueden clasificarse, según el método de selección de los casos, en:

a. Estudios de CC basados en casos.
b. Estudios de CC anidados en una cohorte.
c. Estudios de caso-cohorte.
d. Estudio cruzado de casos.

Estudios de cc basados en casos

Es la estrategia más sencilla para seleccionar los grupos del estudio. En dicha modalidad de estudio, se selecciona a personas con la enfermedad de interés, o con el desenlace de interés como los casos. Pueden seleccionarse casos existentes (por ejemplo, casos de cáncer de esófago ya registrados en el hospital del investigador), y se les llama ***casos prevalentes***. O puede seleccionarse a los pacientes a medida que se los va diagnosticando, y así se los llama ***casos incidentes***. Para los controles, se selecciona a pacientes sin cáncer de esófago, que acudieron (si se toman casos prevalentes) o que acudan (si se toman casos incidentes) al mismo hospital, o provenientes de la misma comunidad que dio origen a los casos, o que pertenezcan a la misma población de referencia.

Debe recalcarse que los controles deben ser representativos de la población que dio origen a los casos, para garantizar la comparabilidad entre los grupos. Es muy importante buscar tal similitud en todos los aspectos relevantes para el estudio: idealmente, deben tener edades similares, el mismo sexo, provenir del mismo estrato socioeconómico, tener hábitos similares y parecerse en los aspectos significativos.

En caso de duda, debe buscarse una fuente distinta para los controles. Cuando se va a seleccionar a pacientes del mismo hospital, debe evaluarse con precaución su origen, así como sus otras características. Y si son personas que provienen de comunidades diferentes o que tienen características de las cuales pueda sospecharse que influirían en el estudio, deberán ser sustituidos por controles que tengan el mismo origen: por ejemplo vecinos o amigos cercanos del caso. También se puede seleccionar a personas que provengan de la misma fuente de referencia; por ejemplo, del mismo centro de salud.

Uno de los sesgos más comunes en esta modalidad de estudio es el ***sesgo de selección***, que ocurre cuando hay desviaciones sistemáticas entre los grupos. Por ejemplo, para un estudio sobre enfermedad hepática por alcohol se toman casos de cirrosis del servicio de gastroenterología del hospital, y como controles, a pacientes del servicio de traumatología de la misma institución. Pero una persona hospitalizada por trauma tiene una mayor probabilidad de ingesta crónica o excesiva de alcohol, lo que alteraría el valor de la asociación que se halle entre alcohol y daño hepático: en ***este*** caso, disminuiría artificialmente y de manera incorrecta esa asociación.

Se asume que tanto los casos como los controles provienen de la misma base de estudio; o sea, que provienen de la misma hipotética cohorte con características similares. Ello es fundamental para asegurar la validez interna del estudio (ver adelante, selección de controles). No obstante debe recordarse que no necesariamente se evitan los sesgos cuando se obtienen

tanto los casos como los controles de la misma base de estudio. Puede ocurrir que haya pérdidas de personas que pudieron pertenecer a los casos o a los controles durante el periodo de inclusión de estos. Por ejemplo, porque murieron las casos más afectados y solo se hallan disponibles para el estudio quienes sobrevivieron (*sesgo de sobrevida*, o de Neyman), o que los controles de mejor condición socioeconómica se hayan desplazado a una nueva clínica abierta en el sector, etc., por lo cual no son tenidos en cuenta cuando se ensamblen los grupos.

Es importante anotar que puede haber diferencias importantes entre casos y controles si unos son tomados durante un periodo prolongado y los otros se toman de casos previos.

Otro posible problema puede ser el *sesgo del recuerdo*, pues en la mayoría de los casos se recolecta la información *después* de que se ha presentado el desenlace, y ello puede llevar a un recuerdo diferencial entre los dos grupos, o a que, simplemente, algunas personas no recuerden sucesos ocurridos hace algún tiempo (meses o años). Como ejemplo, al indagar sobre posibles factores de riesgo para malformaciones congénitas, es muy posible que las madres de niños con anormalidades al nacer tengan más estímulo para recordar, y, efectivamente, se acuerden más de las exposiciones que las madres de niños normales. Esto se podría evitar (si es económicamente posible y si hay un número suficiente de casos incidentes) por medio de la selección concurrente tanto de casos como de controles; o mediante una cuidadosa redacción o la estandarización y la conducción de la entrevista, o aplicando las pruebas o los instrumentos para obtener los datos sobre exposición que averigüen de la manera más neutral posible sobre las exposiciones.

Estudio de cc anidados en una cohorte

Una variedad de estudio de cc es una en la cual el diseño convencional de cc se inserta dentro de una cohorte prospectiva o retrospectiva. Los casos son seleccionados de una cohorte (el término, que hace referencia a un grupo homogéneo muy bien definido en tiempo y espacio, proviene del latín *cohors*, una unidad de la legión romana que se caracterizaba por su uniformidad en formación y entrenamiento; tanto, que, según se decía, actuaban "como un solo hombre"). En este tipo de diseño se selecciona una cohorte que es seguida a lo largo del tiempo, y cada vez que se produzca un caso (desenlace) se selecciona entre las personas que se hallan disponibles en el seguimiento un control (o más de uno; hasta cuatro), al azar. Esto es lo que se denomina *muestreo por densidad de incidencia*, o *muestreo por grupo de riesgo*. Los casos son comparados con un subgrupo de controles obtenidos al azar durante el mismo periodo en el que fueron obtenidos los casos. Ello significa que algunos controles que podrían ser casos en el futuro pueden ser elegidos como controles de los casos ya seleccionados. Esto equivale a hacer un emparejamiento por la duración del seguimiento y permite llevar a cabo procedimientos estadísticos de regresión múltiple para el análisis de datos emparejados y de sobrevida.

En este tipo de estudio se saca provecho de una cohorte existente: por ejemplo, las muestras ya recolectadas en un banco de sangre, cuyos donantes son seguidos durante un tiempo hasta que se presente una enfermedad o un desenlace predefinidos. Cuando se obtiene el número de casos predefinido, se analizan sus muestras de sangre y se comparan con las de una muestra, obtenida en la misma cohorte, de sujetos sin el desenlace (controles) obtenidos durante el mismo periodo en el que se detectan los casos.

EJEMPLO

En un banco de sangre se tienen muestras de plasma recolectadas de un grupo de personas que no han desarrollado úlcera gástrica, y a quienes se les sigue durante un tiempo determinado. Durante el periodo de seguimiento, se les diagnóstica úlcera gástrica a algunos (casos), y cuando se alcanza el cálculo de tamaño de muestra necesario, se analizan sus muestras de sangre y se comparan con las muestras de plasma de sujetos que no hayan desarrollado aún la enfermedad (controles), obtenidos de la misma población en el tiempo durante el cual fueron diagnosticados los casos.

Estudio de caso-cohorte

Este diseño es similar al anterior, ya que se inicia con la construcción de una cohorte de personas o mediante el uso de una cohorte disponible. Durante el seguimiento, se seleccionan los casos incidentes, y una vez se detectan estos, se busca un control seleccionando al azar a una persona

de la cohorte inicial; o sea, de la línea de base. De esta manera, algunos sujetos seleccionados como controles podrían haberse convertido en casos durante el seguimiento. Tal situación obliga a emplear metodologías estadísticas específicas para ese tipo de análisis.

El estudio caso-cohorte tiene la gran ventaja de que si la cohorte al inicio es representativa de la población fuente, podemos obtener la distribución de los factores de riesgo y las prevalencias necesarias para obtener los estimativos del riesgo atribuible poblacional[21]. Así mismo, en esta clase de estudio, el OR provee un estimativo igual al RR.

En el caso de las dos modalidades anteriores, estudios de casos y controles anidados y los estudios de caso-cohorte, en ocasiones se pregunta en ellas por qué no se utiliza toda la población restante como control. Aunque podría hacerse si se cuenta con la totalidad de los datos necesarios (cuestionarios, escalas), en algunos casos deben hacerse pruebas diagnósticas en muestras de suero refrigeradas, y utilizar solo una muestra de los posibles controles significa un ahorro de tiempo y de dinero que hace más eficiente el estudio, y permite guardar algunos sueros para posibles estudios subsecuentes.

A estas dos modalidades (el estudio anidado en una cohorte y el caso cohorte) se les denomina *diseños híbridos*, o ***ambidireccionales***, porque combinan las características y las ventajas de ambos diseños (cohortes, casos y controles). No obstante, los casos son seleccionados utilizando la aproximación al estilo de la modalidad clásica de CC. Las dos modalidades tienen una ventaja especial sobre los estudios de caos y controles basados en casos, ya que permiten reducir el *sesgo de selección.*

Estudio cruzado de casos

Esta clase de estudio, poco común, se lleva a cabo cuando las exposiciones pueden variar a lo largo del tiempo y desencadenar el desenlace. Como ejemplos, la muerte súbita de etiología cardíaca, los intentos de suicidio, las crisis asmáticas, entre otros. En el estudio cruzado de casos, los casos sirven como sus propios controles, de tal manera que el tiempo hasta la aparición del desenlace (por ejemplo, el tiempo inmediatamente anterior a la crisis asmática o al intento de suicido) es tomado como el tiempo para el periodo del caso. Entre tanto, el otro tiempo, el más lejano del episodio (desenlace), se toma como el periodo del control.

Tiene las ventajas de que cada participante es considerado un estrato apareado en el estudio de CC, porque el periodo para el caso es el tiempo previo a la aparición del desenlace, y el periodo para el control, el anterior en la misma persona. Esto permite evitar el sesgo de selección para los controles y elimina, casi del todo, la variabilidad existente entre casos y controles de las otras modalidades. Por otra parte, se hace un autoapareamiento para posibles variables de confusión, como edad, sexo, raza, etc. Tiene, sin embargo, desventajas, como suponer que no hay un efecto de continuación (***carry over***) de la variable de exposición. A la vez, se asume que no hay interacción ni confusión por variables relacionadas con el tiempo, como la temperatura, el día de la semana, etc.; pero dichas variables pueden servir para hacer ajustes o estratificaciones. No obstante, puede darse un ***sesgo de recuerdo***, pues las personas pueden diferir en cuanto a su apreciación del periodo de caso y de control.

Con esta modalidad de diseño puede evaluarse, por ejemplo, el tiempo con buen o mal sueño antes de un accidente automovilístico. Puede definirse el periodo de caso como las 24 horas previas a un accidente, y el periodo de control, como el comprendido entre las 48-25 horas anteriores al accidente. En esta clase de estudio se hace un análisis pareado y una regresión logística condicional.

DEFINICIÓN Y SELECCIÓN DE CASOS

Casos

De acuerdo con Schlesselman (1982)[7], se puede definir un ***caso*** como un sujeto con el desenlace o la enfermedad objeto de estudio. Es vital que la definición de la enfermedad, de la condición de interés o del desenlace sea muy precisa y completa, y que tenga criterios de diagnóstico claros y represente una entidad homogénea. Si no es así, se crearían enormes dificultades con la mala clasificación que se podría dar. Como ejemplo, una definición de demencia de Alzheimer (DA) que se apoye exclusivamente en criterios clínicos del paciente y que no tenga en cuenta los hallazgos psiquiátricos, neuropsicológicos, neurológicos

o de otros exámenes pertinentes (como la resonancia magnética cerebral) es, claramente, una definición incompleta. Al no tomar en cuenta los aspectos omitidos, podrían tenerse muchos diagnósticos errados, que pueden corresponder a otras enfermedades clínicamente similares a la DA, pero en realidad diferentes, como demencias vasculares, demencias frontotemporales o depresión, entre otras. El error diagnóstico puede causar clasificaciones incorrectas, que llevan a un ***sesgo de mala clasificación.***

Por los mencionados inconvenientes, y para evitarlos, se recomienda emplear la definición más precisa, completa y aceptada en el ámbito científico nacional o internacional (por lo general, la CIE-10), así como establecer en el protocolo de investigación criterios precisos de inclusión y de exclusión, de acuerdo con la pertinencia del estudio. Ello permite una caracterización homogénea de la población, aunque ocasionalmente, si son excesivamente estrictos, podrían dificultar que se obtenga la muestra necesaria, por las limitaciones impuestas. Esta dificultad tiene un defecto potencial adicional: disminuye la capacidad para generalizar los resultados, pues estos solo podrán ser aplicados a poblaciones con las características descritas para el estudio. Es lo que se conoce como ***limitación de la validez externa*** (capacidad de generalizar) ***del estudio.***

No obstante lo anterior, en ningún caso se debe privilegiar la validez externa sobre la validez interna, pues sin la segunda no existiría la primera. En esta línea de pensamiento es aconsejable limitar el estudio a los casos en los cuales se pueda obtener información completa y confiable acerca de la enfermedad (desenlace) y en los que se puedan evaluar de manera adecuada la presencia de factores de riesgo o de protección. Ello también aplica para los controles seleccionados.

Ahora, una vez definidos los casos, la siguiente pregunta es: ***¿dónde se los puede encontrar?*** Hay tres fuentes principales para su obtención:

- Casos hospitalarios.
- Casos poblacionales o de la comunidad.
- Casos de grupos especiales.

Casos hospitalarios

Hacen referencia a los casos de personas (pacientes) con la enfermedad que se estudia, diagnosticados o tratados en un determinado hospital o centro de atención, y también en la consulta privada de profesionales de la salud.



Los casos hospitalarios son, por lo general, de fácil acceso, lo cual implica ventajas de tipo económico y logístico. Se puede hacer la escogencia de pacientes por los listados (registros, bases de datos, etc.) que se tengan, y posteriormente, revisar sus historias clínicas, hacer entrevistas o aplicar cuestionarios.

En el primer caso (historia clínica) se corre el riesgo, por demás recurrente, de tener historias clínicas incompletas que no suministran registros confiables de los factores que se van a estudiar, como ya se ha demostrado en la literatura científica[15].

En el segundo caso (entrevistas o aplicación de cuestionarios) se obvia la primera dificultad de información incompleta o imposible de obtener; sin embargo, se incrementan los problemas logísticos, y en algunos casos, también los costos de la investigación; peor aún, se pueden incrementar algunos posibles sesgos característicos de tal tipo de estudios, como sesgos de selección, de recuerdo (memoria) y de información, que serán desarrollados más adelante.

Una tercera opción es tomar a pacientes a medida que se les diagnostica la condición de interés; o sea, casos incidentes. Dicha estrategia prolonga la investigación hasta cuando se logra la muestra determinada, y esa dificultad es aún mayor cuando la enfermedad es rara o de latencia larga. Debe recordarse que en todos los casos seleccionados se debe tener especial precaución con el sesgo del recuerdo (memoria), que se presenta por el hecho de saberse enfermos o con el desenlace definido, lo cual puede llevar a que cuando se recolecten los datos sobre las exposiciones o los factores de riesgo seleccionados, el paciente recuerde con mayor precisión el control o tienda a exagerar lo acontecido, por el simple hecho de tener el desenlace. Por el contrario, puede ocurrir, cuando la enfermedad se encuentra en sus inicios o con pocas manifestaciones, que el paciente tienda a minimizar los recuerdos, lo cual resultaría en una reducción de la probabilidad de detectar factores de riesgo.

Por otra parte, en ambos tipos de consecución de pacientes (prevalentes [antiguos] o incidentes

[nuevos]) se corre el riesgo de cometer un sesgo de selección; esto es, que los pacientes seleccionados, por el hecho de asistir a un determinado centro de salud, pueden no ser representativos de todos los pacientes con la misma enfermedad, pero que no asisten al mismo centro de salud. En tales casos puede disminuir la capacidad para generalizar el estudio, y en ocasiones puede alterar la validez de los resultados, por reducción (o aumento) incorrecto de la medición de la asociación entre el factor y el desenlace.

Casos poblacionales o de la comunidad

Los casos poblacionales se refieren a la localización de los individuos con la enfermedad o el desenlace de interés en una población. Para ello, es necesario tener una lista o un registro de todos los posibles casos, o hacer un censo con el fin de ubicarlos, para posteriormente elegirlos a todos o a una muestra aleatoria de ellos de acuerdo con el tamaño de muestra requerido.

En esta modalidad, de tomar casos poblacionales de un registro de la totalidad de individuos con la enfermedad en una región no es frecuente, pues habitualmente no se conoce a las personas enfermas en la población, y hay evidentes dificultades para detectarlas y localizarlas a todas; en especial, cuando la enfermedad está en sus primeras etapas o es subclínica. Eso ha llevado a que tal tipo de fuente se emplee pocas veces, dados los problemas logísticos y el costo que representa. Sin embargo, cuando se puede realizar tiene ventajas evidentes, como la disminución sustancial de los sesgos de selección, el aumento de la capacidad de generalización y el incremento de la validez interna del estudio.

Casos de grupos especiales

Se habla de un caso de grupo especial cuando se toma a personas (pacientes) con la enfermedad o el desenlace en cuestión que forman parte de un grupo de apoyo de autoayuda, o de otros tan en boga hoy en el campo de la salud, como la sociedad de pacientes con hemofilia, o las asociaciones de pacientes con alzhéimer. Si se van a estudiar rasgos genéticos en pacientes con *delirium tremens*, podría, por ejemplo, solicitarse la colaboración de grupos como Alcohólicos Anónimos.

Recurrir a este tipo de grupos puede facilitar la obtención de casos y de información, pero puede también significar riesgos como los señalados para pacientes hospitalarios; en especial, los sesgos de selección, por cuanto es difícil determinar si quienes asisten a los grupos son semejantes o representativos de quienes no van. ***A priori*, puede pensarse que** hay diferencias entre las personas que solicitan o buscan ayuda y quienes no, o entre los pacientes hipertensos que se afilian a una institución especializada para su tratamiento y quienes no buscan este tipo de ayuda. Por otra parte, en las enfermedades de alta mortalidad podemos llegar a conclusiones que hagan referencia tan solo a personas sobrevivientes (***sesgo de supervivencia*).**

Dados estos posibles problemas de selección, debemos utilizar dichos grupos con las limitaciones que les son propias y proveer alternativas para disminuir las dificultades inherentes a cada proyecto específico.

Principios de comparabilidad y selección de controles

Por lo común, no se le presta la suficiente atención a la selección de controles. Se emplea poco tiempo en estudiar las dificultades que pueden surgir de su selección y las implicaciones que conllevan para el estudio. Esta es, sin embargo, la fase cardinal para dicho diseño, y cuando se hace correctamente es la principal garantía de la validez de los resultados y de las conclusiones obtenidas.

Se han dado muchas definiciones de controles. Una de las más claras define al control como el individuo que podría haber sido seleccionado como caso si hubiera desarrollado el desenlace o la enfermedad en estudio.

En su libro sobre diseños de cc, Schlesselman[7] describe que un control es un individuo en quien la condición o la enfermedad de interés no están presentes, y a quien se utiliza para compararlo con los casos.

Para seleccionar al grupo de control, Wacholder ***et al.* (1992) describieron cuatro principios** de comparabilidad[22,23] **que deben ser tenidos en** cuenta en la selección:

a. **Principio de la base del estudio.**
b. **Principio de control de la confusión (en** inglés, ***deconfounding***).
c. **Principio de precisión comparable.**
d. **Principio de eficiencia.**

Principio de la base del estudio

Los CC ocurren dentro de una determinada cohorte. Esta puede ser abierta (es decir, no bien definida) en la mayoría de estudios de CC basados en casos, o cerrada (en los diseños de caso cohorte o CC anidados en una cohorte, en los cuales los controles provienen de un grupo mejor definido, y que, a su vez, dio origen a los casos). Esa población de referencia es la que se denomina ***base de estudio***.

Tal denominación se refiere a los miembros pertenecientes a una cohorte, o segmento de población que dio origen a los casos (sujetos enfermos o con el desenlace) para el periodo cuando fueron elegidos como casos[24,25]. La clarificación de la base de estudio se reduce o elimina el sesgo de selección. Cabe anotar que la base de estudio puede ser ***dinámica***; esto es, que en ciertos momentos la persona hace parte de ella, y en otros, no.

La base de estudio se divide en ***primaria*** y *secundaria*[26]. La primaria se refiere al grupo que el investigador quiere estudiar, en una población temporal y geográficamente definida, lo que hace necesarios sistemas para la identificación de todos los casos, excepto cuando se trata de una cohorte cerrada. El problema principal con los estudios de CC con base primaria reside, entonces, en la identificación de todos los casos; o sea, en tener una cohorte definida y con activo reconocimiento de los casos. Si tales condiciones se logran, una vez detectados los casos se procede a identificar a los controles de la misma base primaria; idealmente, haciendo la selección mediante alguna técnica de muestreo aleatorio **(figura 16.2)**.

La base primaria puede traer beneficios para la selección y el muestreo de los controles, aun cuando trae inconvenientes en la detección de casos leves u otros que, por circunstancias especiales, no son detectables[27,28]. Por ejemplo, si se tiene una población espacial y temporalmente definida y se quiere estudiar la presencia de signos leves de depresión como desenlace, es muy probable que muchos de ellos no sean detectados, o lo sean difícilmente en una búsqueda activa. Ello podría promover un ***sesgo de mala clasificación*** del desenlace, que podría mover el resultado en el sentido de la hipótesis nula (es decir, el resultado se desviaría hacia la no diferencia entre los grupos).

La base secundaria hace referencia a que los casos son identificados antes de que la base de estudio se encuentre claramente identificada; por ejemplo, al seleccionar los casos en alguna institución de salud. Ello ocurre, generalmente, en los estudios de CC basados en caso. En dichos estudios (de hecho, la mayoría de los estudios de CC reportados en la literatura), la base secundaria definida secundariamente es la fuente de los casos (sujetos enfermos o con el desenlace), y los controles se seleccionan como pertenecientes

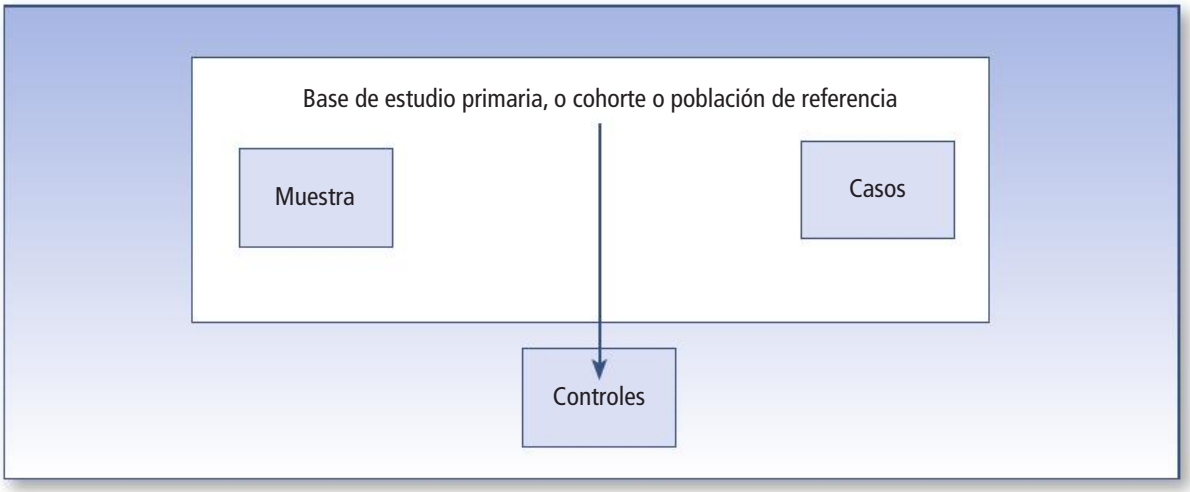


FIGURA 16.2 Selección de controles de la base primaria de estudio.

a dicha base de estudio, teniendo en cuenta que estos podrían haber sido detectados como casos en el evento de que hubieran enfermado durante el tiempo de la investigación.

El mayor problema para los mencionados estudios es, entonces, la definición de la base secundaria, pues ella provee el soporte para que en el estudio se dé una adecuada selección de los controles. A diferencia de la base primaria, en la que se parte de la cohorte o la población de referencia definida, en esta se parte de los casos seleccionados y se tienen en cuenta su procedencia y sus características sociodemográficas para construir la base de la cual, posiblemente, emanaron, para de allí seleccionar los controles, o, según el protocolo, seleccionarlos de la institución de salud, siempre teniendo en cuenta los patrones de referencia y características similares a los que se tuvieron para los casos. El objetivo final es lograr la mayor comparabilidad con el grupo de casos.

Los estudios con base en casos que se inician con pacientes de un hospital, de donde secundariamente se construye la base secundaria, parten del supuesto de que los controles hospitalarios tienen la misma distribución de factores de exposición y riesgo que en la base primaria. Dicho supuesto, sin embargo, no se cumple en todas las circunstancias, pues se requeriría que tanto los casos como los controles hubieran sido admitidos al mismo hospital en el evento de haber sufrido la otra enfermedad (los casos, haber sufrido la enfermedad que llevó a los controles a ser admitidos, o los controles, haber sufrido el desenlace de interés). Se requiere, además, que la admisión del sujeto "caso" y del "control" sean independientes del factor o los factores de exposición[28].

La atención a la base de estudio y a la necesidad de obtener los controles en la misma base que da origen a los casos es fundamental, pues tiende a disminuir el sesgo de selección, muy probable en este tipo de estudios.

Principio de control de la confusión

Este principio (en inglés, ***deconfounding***) se refiere a la necesidad de tomar en cuenta todos los posibles factores de confusión presentes en el estudio que pueden sesgar o distorsionar los resultados de la investigación. En la fase de protocolo del estudio se deben determinar los potenciales factores de confusión (factores asociados a la exposición que se estudia, también asociados al desenlace objeto de estudio, y que no pertenezcan a la cadena causal) presentes en el estudio, y debe planearse el control durante el análisis de los datos, bien sea mediante análisis estratificado, o bien, mediante modelos multivariados. En caso de que no sean medidos, los potenciales factores de confusión pueden, efectivamente, hacer ininteligibles y comprometer los resultados finales.

EJEMPLO

Si se quieren evaluar factores de riesgo para la aparición de cáncer de lengua (desenlace), debe tenerse en cuenta que el consumo de alcohol puede figurar con una asociación alta, pero al analizar el consumo de cigarrillo o de tabaco se verá que el alcohol pierde peso como factor (factor de confusión), y se puede confundir la asociación causal de los dos primeros. La selección de los controles debe buscar, en todo momento, minimizar la posibilidad de confusión.

Principio de precisión comparable

Este principio tiene que ver con la necesidad de evaluar y de medir la exposición o el factor de riesgo o de protección que interesa al estudio, con los mismos instrumentos y de manera similar y comparable en ambos grupos (casos y controles). La evaluación diferencial o con mayor énfasis (generalmente, en el grupo de casos) puede llevar a un sesgo de información que pone en riesgo la comparabilidad de los dos grupos y la validez de los resultados del estudio. Al garantizar que se miden las exposiciones de la misma manera, se evitan los sesgos y se garantiza que los errores de medición que pueden ocurrir serán similares entre los dos grupos o no serán ***errores diferenciales***, expresión con la que se designa el predominio de error en uno de los dos grupos.

Principio de eficiencia

Se refiere a la necesidad de toda investigación de obtener en el menor tiempo posible y con recursos limitados el máximo de información y las mejores conclusiones. Cuando se seleccionan los controles se debe tener en cuenta los costos para obtenerlos, así como la facilidad, validez y precisión de las medidas utilizadas.

La eficiencia se refiere a dos aspectos:

a. La ***eficiencia estadística***, que se relaciona con el poder del estudio (1-error β), y con el em-

parejamiento, cuando este se hace necesario (en pocas situaciones).

b. **La *eficiencia operacional*, que se refiere a** la necesidad de tener en cuenta el tiempo, la energía y los recursos financieros para implementar una determinada modalidad de selección de controles, así como a la justificación de tener varios controles por caso. Si bien se logra una mejor precisión al tener más de un control por caso, se ha estudiado y descrito que a partir de cuatro controles por caso se logra muy poco en cuanto a precisión, y los costos y los procesos aumentan de manera desproporcionada. Por tal razón, se recomienda tener, como máximo, tres controles por caso.

Elección de controles

Se deben tener en cuenta los cuatro principios mencionados para la selección de los controles. Adicionalmente, para seleccionar controles deben tenerse en cuenta las siguientes características:

- El control es un sujeto (persona) que pertenece a una ***muestra no sesgada* de personas, que** no ha desarrollado la enfermedad en estudio y puede estar o no expuesto (***independencia***) a los factores de riesgo o de protección bajo estudio.

El control debe ser representativo de las personas que no han desarrollado la enfermedad y pertenecen a la población (*base de estudio* primaria o secundaria) que dio origen a los casos.
Como ya se mencionó, para que un control sea adecuado debería haber podido ser escogido como caso si hubiera desarrollado la enfermedad o el desenlace en estudio.

- El control debe seleccionarse (al igual que los casos) siempre de manera independiente de la exposición al potencial factor de riesgo (o de protección) en estudio. Ello es muy importante: para la selección del caso o del control, solo debe tenerse en cuenta la presencia o la ausencia del desenlace de interés, y no se debe conocer si está o ha estado expuesto.

Fuentes de controles

Se han descrito múltiples fuentes de controles, las cuales deben cumplir las condiciones ya señaladas. Cada una de dichas fuentes tiene aspectos positivos y negativos que deben conocerse:

Controles obtenidos de sujetos hospitalizados o en tratamiento

Este grupo hace referencia a sujetos con una enfermedad diferente de la que es objeto de estudio. Como ejemplos, para casos con demencia frontotemporal, los controles pueden ser pacientes con enfermedad diverticular no complicada, carcinoma de páncreas, conjuntivitis crónica, etc., que cumplan con las características anotadas en el aparte de comparabilidad.

Las ventajas de dichos controles son la mayor facilidad para conseguirlos (en la institución de salud o en el hospital), su mayor probabilidad de hallarse disponibles para colaboración con la entrevista o los cuestionarios que se aplican (en tal sentido, los enfermos tomados como controles tienden a brindar mayor información), su menor costo (por la facilidad de obtenerlos) y su menor probabilidad de sesgo de recuerdo (los enfermos tienden a recordar mejor, con más detalle y de manera más exhaustiva).

Como desventajas, podrían ser menos representativos de la población general, por no pertenecer a la misma base de estudio que los casos; en particular, cuando los controles (que tienen otra enfermedad) pudieron haber seguido un proceso de remisión o de admisión diferente del de los controles.

Controles obtenidos de la población general

Este grupo se puede obtener a partir de listados de la población (listas de centros de salud, de entidades promotoras de salud [EPS], de las alcaldías, de las oficinas nacionales o regionales de estadística, etc.) o mediante contactos cara a cara, o llamadas telefónicas al azar, convocatorias por redes sociales, etc. La estrategia tiene como ventaja que si se obtienen de la misma base de estudio que dio origen a los casos, se logra disminuir la probabilidad de sesgo de selección. Sin embargo, tiene varias desventajas en comparación con el grupo de controles anterior, pues requieren mayor presupuesto y tiempo, son más difíciles de conseguir, los sujetos involucrados tienen, por lo general, menor motivación para colaborar y responder lo requerido, y tienen menor disposición para recordar situaciones

específicas, por no estar enfermos (sesgo del recuerdo).

Controles obtenidos de grupos especiales

Esta fuente de controles hace alusión a amigos de la persona seleccionada como caso, familiares de este, la pareja del enfermo (caso) u otros grupos que, por su existencia, sean fáciles de obtener (asociaciones de pacientes o de otro tipo). Tienen como ventaja una mayor colaboración con el estudio, por su compromiso con el sujeto escogido como caso o con la sociedad; pero la desventaja que pueden tener es el posible sobreemparejamiento de los grupos (pueden volverse tan parecidos, por tener hábitos similares, que termina no habiendo diferencias entre ellos).

Podría pasar que si el control se selecciona entre los amigos del paciente caso, sus hábitos y sus características de estilo de vida sean muy similares, y podría pasar que se pierda la asociación buscada. Al investigar si ser sedentario es un factor de riesgo para tener infartos, es posible que los amigos del caso tengan también sus hábitos sedentarios, por lo cual se perdería la opción de encontrar que el ejercicio puede prevenir o disminuir el riesgo de infartos en el corazón.

La selección del control debe ser muy cuidadosa en este tipo de estrategia, y si bien se debe buscar que tengan alguna relación con el caso, debe vigilarse que la relación no sea tan cercana como para alterar la importancia de la exposición. Si se seleccionan los controles para un caso de diabetes con infarto cerebral entre los pacientes de alguna asociación de diabéticos, sus hábitos de vida, ejercicio, dieta y otras características pueden ser muy similares a las del caso, y no ser representativas de la población general.

Tanto la capacidad de recuerdo como las características que impone la pertenencia a un grupo (por ejemplo, la representatividad de la base de estudio) son algunos de los aspectos que se deben tener en cuenta para seleccionar al grupo control. Si se tienen dudas sobre la idoneidad del grupo seleccionado, queda la opción de obtener dos o más grupos controles, tomados de fuentes distintas. Si los resultados obtenidos son similares en las medidas de asociación cuando se los contraste con el grupo de casos, podrá aceptarse que los resultados son válidos y que los grupos controles fueron bien seleccionados.

SESGOS

Una de las principales dificultades de este diseño es la posibilidad de tener errores sistemáticos (sesgos) en algunas de las fases del estudio[29]. Algunos de los descritos más a menudo son:

- Sesgo de recuerdo.
- Sesgo de selección.
- Sesgo de medición.
- Sesgo de mala clasificación.

Sesgo de recuerdo

Se define como los errores sistemáticos que puede haber en el recuerdo o en el reporte de la información sobre exposición a factores de riesgo o de protección en los casos o en los controles, y que pueden suscitar cambios en las verdaderas diferencias entre los grupos. A este respecto, se sabe que quienes tienen una enfermedad tienden a buscarle explicaciones, lo cual suscita cavilaciones y teorías sobre su origen. Eso puede llevar a que se les dé importancia (en ocasiones, inusitada) a detalles o a recuerdos que habrían pasado inadvertidos si la persona no hubiera estado enferma.

EJEMPLO

Un estudio de cc que busca factores de riesgo para la ocurrencia de defectos del tubo neural en los hijos de pacientes con trastorno bipolar (casos) toma como controles a hijos de pacientes con diabetes gestacional. Al indagar por experiencias y ocurrencias durante las etapas anteriores al parto y a lo largo de la niñez temprana, es muy probable que las madres con trastorno bipolar y sus familiares recuerden más detalles acerca de aspectos previos al embarazo que los controles o sus familiares, pues la diabetes gestacional debuta durante el embarazo, lo cual los lleva a estar menos alerta sobre sucesos anteriores a este; por el contrario, los sucesos que se den durante el embarazo (por ejemplo, la toma de medicamentos) tendrán la misma probabilidad de recordación en ambos grupos.

Sesgo de selección

Se produce cuando la inclusión de los casos o de los controles depende, en cierta medida, de la exposición de interés (factor de riesgo o de protección). En el clásico estudio sobre el síndrome de choque séptico y el uso de tampones, el efecto

publicitario acerca de la posible asociación entre el uso de tampones y la presencia de infecciones que llevarían a un choque séptico llevó a muchas mujeres que usaban tampones a consultar a su médico por cualquier síntoma o signo similar a los reportados en los medios de comunicación. La probabilidad del diagnóstico aumentó, por cuanto el profesional tendía a diagnosticarlo más a menudo, por el efecto negativo que podría tener su no detección (falsos positivos). Dichas circunstancias llevaron a un aumento artificial en la asociación entre la entidad estudiada y el uso del tampón[30].

Puede ocurrir también un sesgo de selección en los casos donde se paga por la participación en estudios clínicos. Como los casos se autoseleccionan, aunque cumplan los criterios de inclusión y de exclusión, puede ocurrir que quienes buscan participar puedan tener con más probabilidad exposiciones determinadas que se pueden asociar al desenlace. Si un individuo tiene bajos recursos (y se ve así más estimulado a participar en el estudio, por cuanto obtendrá dinero por su participación), puede tener mayor posibilidad de padecer enfermedades pulmonares por hacinamiento, lo cual no sería conveniente en un estudio que evalúe dicha asociación. Por otra parte, es indispensable estar atento a los sesgos de supervivencia que pueda haber en los casos, los cuales pueden llevar a que se seleccione a los casos menos graves, que tengan diferencias significativas con quienes ya murieron en cuanto a sus niveles de exposición a los factores de interés.

Sesgo de medición

Se refiere a la situación en la cual los investigadores o los encuestadores hacen un más detallado y profundo interrogatorio o encuesta a alguno de los dos grupos (casos o controles). En ciertas situaciones, y en medio del afán por comprender adecuadamente los posibles factores de riesgo o los factores asociados a cierta condición, se puede hacer más énfasis en la aplicación de los sistemas de medición en el grupo de casos. Para prevenir esa dificultad, es conveniente que los encuestadores o las personas encargadas de recolectar la información sobre las variables de interés del estudio permanezcan enmascarados a las hipótesis del estudio, e, idealmente, no sepan a cuál grupo pertenece la persona encuestada.

Es fundamental el entrenamiento previo para los encuestadores, para que puedan recolectar la información de manera estandarizada y similar en todos los sujetos. En algunos casos, se pueden repetir, al azar, cierto número de entrevistas y el proceso de obtención de datos en los sitios definidos para ello (historia clínica, bases de datos, etc.), y tanto para casos como para controles, como una buena opción para verificar la calidad de los datos.

Sesgo de mala clasificación

Se refiere a errores en la categorización de la enfermedad o de la exposición, como catalogar en el grupo control a un caso, o viceversa; o el error de decidir que se tiene la exposición cuando, en realidad, es lo contrario. Henneckens y Buring (1987) describen que las consecuencias de la inadecuada clasificación dependen de si la mala clasificación en un eje (exposición o enfermedad) es o no independiente de la clasificación en el otro eje. La mala clasificación será ***aleatoria*** o *no diferencial* cuando la proporción de clasificaciones incorrectas (como sujetos enfermos o expuestos) es similar en ambos grupos y el porcentaje de error es similar tanto en los casos como en los controles. Este tipo de sesgo lleva, por lo general, a subestimar el RRI; esto es, a un *sesgo hacia la hipótesis nula de no diferencia*. Si dicho porcentaje de error es diferente en ambos grupos, el sesgo de mala clasificación sería diferencial o no aleatorio, y, según la situación, puede incrementar o disminuir el RRI u OR.

RAZÓN DE CONTROLES A CASOS

La mayoría de las investigaciones sobre estudios de CC tienen una proporción de un control por cada caso. Sin embargo, cuando se tiene un número reducido de casos (y, con ello, una disminución del poder del estudio), se puede incrementar la cifra aumentando el número de controles por caso hasta un límite máximo de tres a cuatro controles por caso. A partir de dicho número, se hace ineficiente aumentar los controles, pues se gana poco y la obtención del control puede significar dificultades logísticas o costos, y habrá muy poca ganancia en cuanto a poder[31]. En la mayoría de situaciones, ante la falta de poder, debe optarse por una relación de uno a dos o a tres controles por caso, según la necesidad del estudio **(figura 16.3)**.

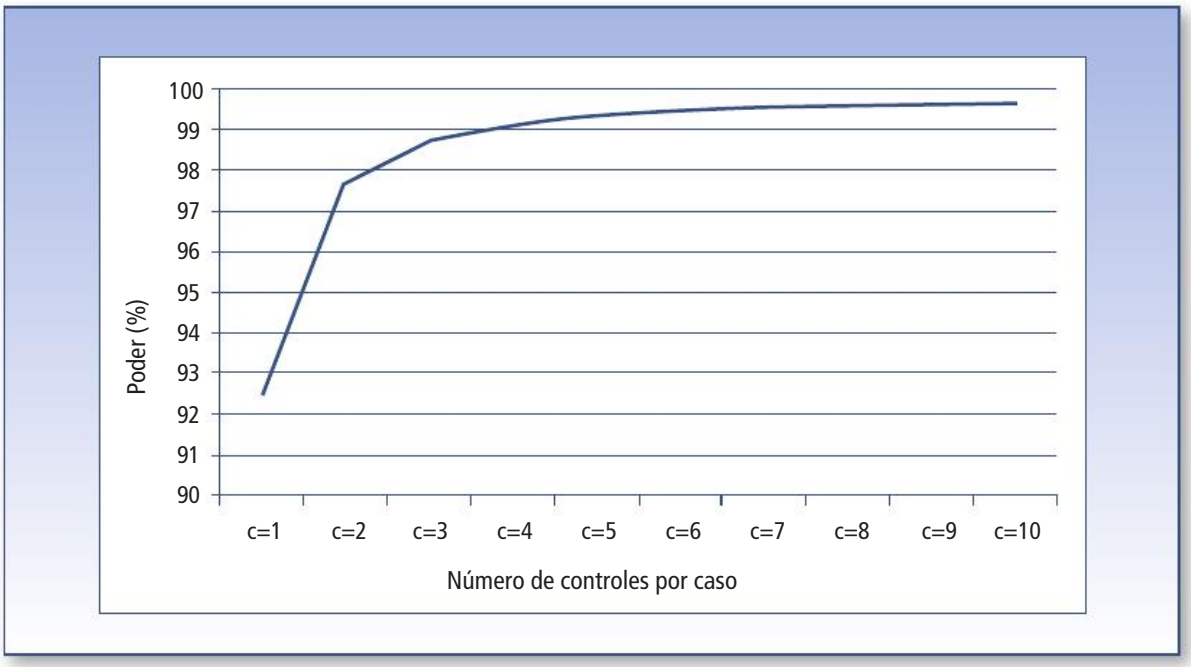


Figura 16.3 **Poder según el número de controles por caso, al considerar los siguientes parámetros: OR de 2, con prevalencia del** 50 %, con un nivel de significación del 5 %, a 2 colas, y tamaño de la muestra de 200 casos.

Si es posible, resulta una opción mucho mejor aumentar el número de casos (pasar de 135 a 165 casos, por ejemplo), en vez de aumentar el número de controles por caso. A veces, puede aumentarse la base poblacional (geográfica) o temporal del estudio, y al aumentar el número de casos se obtendrá un incremento en la precisión del estudio mayor que cuando se amplía el número de controles por caso.

Emparejamiento

En medicina, hay la creencia de que el emparejamiento (***matching***) **es una estrategia** deseable y beneficiosa para los estudios de **CC**, y que refleja una mejor calidad del diseño. Tal apreciación no tiene mucho sentido; el apareamiento puede llevar, por el contrario, a incrementar las dificultades logísticas y los costos, y a dificultar y complicar el análisis estadístico y la interpretación de los resultados.

Aunque intentar parear por ciertas variables (sexo, edad, raza, etc.) suena, en ocasiones, aconsejable, se corre el riesgo de sobreemparejamiento de los grupos, con una disminución o una pérdida de la opción de detectar diferencias, si las hay, pues en la medida en que se hagan más similares los grupos respecto a una variable (por ejemplo, sexo) o a un grupo de variables (por ejemplo, sexo y nivel educativo), puede aumentar el riesgo de sesgos, y disminuir la eficiencia del estudio[32].

Por lo anterior, se recomienda el emparejamiento tan solo en los siguientes casos[33]:

- Cuando un factor de confusión es un factor de riesgo muy grande para la enfermedad (**OR** > 7). Por ejemplo, tabaquismo como factor de confusión de la relación que se estudia entre abuso del alcohol y cáncer de pulmón.
- Cuando la exposición es muy rara.
- Cuando la muestra es muy pequeña.
- En todos los casos donde se hace emparejamiento es obligatorio un análisis pareado (por ejemplo, utilizar en el análisis bivariado la prueba de McNemar, en vez de la prueba de χ^2), así como estratificar en el análisis o utilizar métodos estadísticos pertinentes al emparejamiento.

VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS DE CC

Este diseño, uno de los más utilizados en la medicina clínica, tiene varias ventajas:

- En promedio, se hace en menos tiempo que otros estudios; sobre todo, en tiempos

menores que los de estudios de cohortes concurrentes o prospectivas.
- En general, es menos costoso que otros estudios, por su relativa facilidad y por su rapidez para la obtención de casos.
- Sirve para evaluar enfermedades poco comunes o raras.
- Examina simultáneamente múltiples factores de riesgo o de protección para una enfermedad o un desenlace estudiado.
- Es conveniente para estudiar enfermedades de larga latencia.

Este diseño tiene, sin embargo, algunas desventajas:
- Es poco eficiente para evaluar exposiciones raras.
- No permite determinar frecuencias.
- No puede calcular tasas de incidencia de enfermedad en sujetos expuestos y no expuestos (excepto cuando se basa en poblaciones).
- En ocasiones, es difícil establecer la relación temporal entre exposición y enfermedad.
- Es susceptible a diversos sesgos, dentro de los cuales son más notorios los sesgos de selección y de recuerdo (memoria).

MEDIDA DE ASOCIACIÓN Y ANÁLISIS

Una vez recolectada la información, tanto para el grupo de casos como para el de controles, sobre la exposición a los factores de riesgo o de protección de interés, se calcula la medida de asociación pertinente a este tipo de estudios, el **RRI (también llamado razón de suertes o razón** de momios). El **RRI** describe la magnitud de la asociación entre la exposición y la enfermedad o el desenlace estudiado; o sea, cuántas veces más riesgo tienen los casos de haber estado expuestos, en comparación con los controles. Así, en un estudio sobre la asociación entre el cáncer de pulmón y el hecho de haber estado expuesto al cigarrillo se obtiene un **RRI de 16, lo** cual indica que es 16 veces más probable haber estado expuesto al cigarrillo si se tiene cáncer pulmonar que si no se tiene dicho tumor.

Debe tenerse la precaución de interpretarlo de la manera descrita, y no caer en el error de interpretarlo como un riesgo relativo directo. A partir de un estudio experimental, un **RR = 3,0** debe interpretarse como que la exposición triplica el riesgo del desenlace. En un estudio de casos y controles, un **RRI = 3,0 debe interpretarse como** que es tres veces más probable que el caso haya estado expuesto, en comparación con el control.

El **RRI se calcula a partir de una tabla de** 2 × 2, de la siguiente manera **(figura 16.4).**

	Casos	Controles
Expuestos	A	B
No expuestos	C	D

$$RRI = \frac{A \times D}{C \times C}$$

FIGURA 16.4 Tabla de 2 x 2 para calcular el RRI.

EJEMPLO

En un estudio que busca determinar la asociación entre la exposición a un determinado fármaco durante el embarazo y la presencia o no de labio leporino u otra anomalía congénita en los hijos, se encuentran los siguientes resultados **(figura 16.5)**:

Fármaco	Casos (anomalías presentes)	Controles (sin anomalías)
Expuestos	50	30
No expuestos	10	35

$$RRI = \frac{50 \times 35}{30 \times 10} = 5{,}8$$

FIGURA 16.5 Resultados de asociación.

En el estudio descrito, el **RRI** es de 5,8, lo cual se interpreta así: las madres de hijos con anomalías congénitas tienen un riesgo 5,8 veces mayor de haber estado expuestas al fármaco que las madres de hijos sin anomalías congénitas.

Ahora bien, el citado hallazgo puede explicarse por algunas variables de confusión que pueden explicar la relación encontrada. Algunas de ellas pueden ser la edad de las madres, el consumo de sustancias psicoactivas durante la adolescencia, laborar en una empresa que utiliza fungicidas, etc. Si tomamos la edad de las ma-

dres, con la idea de que las madres mayores de 35 años ("madres añosas") tienen mayor riesgo de concebir a hijos con anomalías que quienes no lo son, podemos hacer lo que se denomina un *análisis estratificado*, y así obtener dos tablas de 2 × 2: una para madres menores de 35 años, y otra para madres de 35 años o mayores **(figuras 16.6 y 16.7)**:

Tóxico	Casos (madres de hijos con anomalías)	Controles (madres de hijos sin anomalías)
Expuestos	15	20
No expuestos	6	8

$$\text{(Riesgo relativo indirecto (RRI)} = \frac{15 \times 8}{6 \times 20} = 1$$

FIGURA 16.6 Madres menores de 35 años.

Tóxico	Casos (anomalías)	Controles (sin anomalías)
Expuestos	35	10
No expuestos	4	27

$$\text{Riesgo relativo indirecto (RRI)} = \frac{35 \times 27}{4 \times 10} = 23{,}6$$

FIGURA 16.7 Madres de 35 años o mayores.

Al controlar por la variable edad, se hace evidente que las madres que dieron a luz a hijos con anomalías congénitas tienen 23,6 más riesgo en el grupo de las madres mayores de 35 o más años y expuestas al fármaco que las madres menores de 35 años expuestas al fármaco (RRI = 1). Ello puede sugerir una interacción entre la edad de la madre y la exposición a fármacos.

Este ejemplo muestra la necesidad de buscar los posibles factores de confusión y de analizar los datos mediante un análisis estratificado de las posibles variables de confusión estudiadas.

CONCLUSIÓN

El estudio de CC es de gran utilidad para la búsqueda de factores de riesgo o de protección asociados a un desenlace (enfermedad) en estudios de ciencias de la salud. Para hacer tal tipo de estudio, debe definirse con cuidado qué es un caso y qué es un control. Dicha definición es fundamental para el estudio: sobre el tipo de CC que se tengan a mano, se definirá la generalizabilidad del estudio. Los estudios de CC tienen ventajas muy bien definidas, como su mayor rapidez, su menor costo y su gran utilidad para evaluar factores de riesgo y de protección asociados a enfermedades raras. Su mayor punto débil es la susceptibilidad a sesgos de diversa índole, como sesgos de selección, de mala clasificación y de recuerdo (memoria).

Referencias

1. Lilienfeld AM, Lilienfeld DE. A century of case-control studies: progress? J Chronic Dis. 1979;32:5-13.
2. Lilienfeld DE. Foundations of epidemiology. 3rd ed. New York: Oxford University Press; 1994.
3. Cole P. The evolving case-control study. J Chronic Dis. 1979;32:15-27.
4. Broders AC. Squamous-cell epithelioma of the lip: a study of five hundred and thirty-seven cases. J Am Med Assoc. 1920;74:656-64.
5. Lane-Claypon J. A further report on cancer of the breast: reports on public health and medical subjects. Londres: Majesty Stationery Office; 1926.
6. Lombard HL, Doering CR. Cancer studies in Massachusetts: Habits, characteristics and environment of individuals with and without cancer. N Engl J Med. 1928;198:481-7.
7. Schlesselman J. Case-control studies: Design, conduct, analysis. New York: Oxford University Press; 1982.
8. Proctor R. Angel H Roffo: the forgotten father of experimental tobacco carcinogenesis. Bull World Health Organ. 2006;84:494-6.
9. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. J Am Med Assoc. 1950;143:329-36.

10. Levin ML, Goldstein H, Gerhardt PR. Cancer and tobacco smoking; a preliminary report. J Am Med Assoc. 1950;143:336-8.
11. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. Br Med J. 1950;2:739-48.
12. Armenian HK, Lilienfeld DE. Applications of the case-control method. Overview and historical perspective. Epidemiol Rev. 1994;16:1-5.
13. Gómez C. Introducción al estudio de casos y controles. Rev Colomb Neumol. 1997;10:132-5.
14. Slone D, Shapiro S, Rosenberg L, et al. Relation of cigarette smoking to myocardial infarction in young women. N Engl J Med. 1978;298:1273-6.
15. Fonseca MG, Armenian HK. Use of the case-control method in outbreak investigations. Am J Epidemiol. 1991;133:748-52.
16. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst. 1951;11:1269-75.
17. Kish L. Some statistical problems in research design. Am Sociol Rev. 1959;24:328.
18. Berkson J. The statistical study of association between smoking and lung cancer. Proc Staff Meet Mayo Clin. 1955;30:319-48.
19. Lilienfeld AM, Graham S. Validity of determining circumcision status by questionnaire as related to epidemiological studies of cancer of the cervix. J Natl Cancer Inst. 1958;21:713-20.
20. Hennekens CH. Epidemiology in medicine. 1st ed. Boston: Little, Brown; 1987.
21. Szklo M. Epidemiology: beyond the basics. 3rd ed. Sudbury, Mass: Jones and Bartlett Publishers; 2012.
22. Lasky T, Stolley PD. Selection of cases and controls. Epidemiol Rev. 1994;16:6-17.
23. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992;135:1019-28.
24. Miettinen OS. The "case-control" study: valid selection of subjects. J Chronic Dis. 1985;38:543-8.
25. Miettinen O. Response. J Clin Epidemiol. 1986;39:567.
26. Poole C. "Would" vs "should" in the definition of secondary study base. J Clin Epidemiol. 1990;43:1016-20.
27. Savitz DA, Pearce N. Control selection with incomplete case ascertainment. Am J Epidemiol. 1988;127:1109-17.
28. Gómez C, Ardila J. Estudio de casos y controles. En: Ruiz A, editor. Epidemiología clínica investigación clínica aplicada. Bogotá: Médica Panamericana; 2001. p. 265-78.
29. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32:51-63.
30. Stallones RA. A review of the epidemiologic studies of toxic shock syndrome. Ann Intern Med. 1982;96(6 Pt 2):917-20.
31. Ury HK. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. Biometrics. 1975;31:643-9.
32. Breslow N. Design and analysis of case-control studies. Annu Rev Public Health. 1982;3:29-54.
33. Evans A, Lozano J. Advanced topics in clinical epidemiology. Confounding and interaction. Bogotá: Pontificia Universidad Javeriana, Departamento de Epidemiología Clínica y Bioestadística; 1995.

CAPÍTULO

17 Estudios de cohorte analítica

JUAN GABRIEL RUIZ PELÁEZ

INTRODUCCIÓN

Aspectos generales

En su forma más simple, una ***hipótesis de asociación*** propone que la exposición a un factor "A", (***variable de exposición de interés***, o independiente) se asocia a un cambio en la frecuencia o la intensidad de presentación de otra variable "B" (***variable de desenlace***, resultado o dependiente). Se conjetura que la frecuencia de presentación, o la magnitud de la variable "B", es diferente en sujetos expuestos al factor de interés "A" cuando se compara con sujetos no expuestos ("no A"), o expuestos a otro factor "C" (el ***comparador***, o exposición de control).

Si hay otros factores diferentes de la exposición de interés (otras variables independientes, "X", "Y" y "Z") que también se asocien a cambios en la frecuencia o la intensidad del desenlace "B", su presencia podría explicar las diferencias en la frecuencia o la magnitud de "B". La hipótesis debe, entonces, especificarse de tal manera que controle la confusión o la modificación de efecto atribuible a esas otras variables "X", "Y" y "Z".

La hipótesis podría formularse así: en un grupo de sujetos susceptibles de padecer "B", pero libres de esa condición al comienzo del estudio, los individuos expuestos a "A", al cabo de cierto intervalo "T" van a presentar "B" con una frecuencia diferente de la de los sujetos expuestos a "C", siempre y cuando la distribución de los factores "X", "Y" y "Z" en los grupos expuestos a "A" y a "C" esté balanceada, y no haya interacciones entre ninguna de las exposiciones "A" o "C" y cualquiera de los otros factores "X", "Y" y "Z".

La conducción de un experimento suele ser la manera más natural y eficiente de evaluar este tipo de hipótesis. En un experimento, la exposición a los factores de interés es decidida por el investigador, para satisfacer las necesidades del estudio (idealmente, de forma aleatoria, para prevenir sesgos en la asignación), y no en atención a las necesidades del sujeto participante.

El investigador ensambla un grupo de sujetos libres de la condición "B", pero que tienen la posibilidad (***susceptibilidad***) de padecer "B", y los asigna a exponerse a "A", o a "C". La manera como los asigne a "A" y a "C" debe permitir que los factores "X", "Y" y "Z" se distribuyan de forma balanceada (no diferencial) entre expuestos a "A" y a "C". La razón por la cual la asignación aleatoria es la mejor manera de hacer dicha asignación se discute en el capítulo sobre experimentos clínicos. En esencia, el argumento principal a favor de asignar aleatoriamente es

que permite hacer una estimación no sesgada de la asociación entre "A" y "B", y, adicionalmente, tiende a balancear la distribución de potenciales modificadores de efecto y factores de confusión tales como "X", "Y" y "Z".

En la investigación en salud en seres humanos, hay muchas circunstancias en las que la exposición experimental a potenciales factores de riesgo, pronóstico o causantes de desenlaces desfavorables (enfermedad, disfunción, dolor, muerte, etc.) es éticamente inaceptable. En otros casos, aun cuando no hubiese objeciones éticas, puede, simplemente, no ser factible hacer una asignación experimental, como cuando el factor estudiado es un ***atributo*** (sexo, raza, grupo sanguíneo) que el investigador no puede modificar, sino que debe limitarse a documentar (***asignación observacional***). En tales casos, la única forma viable de estudiar asociación y causalidad es a través de ***estrategias observacionales***.

Como ya se mencionó, en un experimento aleatorio se hace una estimación no sesgada de la asociación entre exposición y desenlace. Por una parte, se estima la frecuencia esperada del desenlace "B" en relación con "A", para todos los participantes en el estudio, mediante la medición de la frecuencia observada de "B" en una muestra representativa (escogida aleatoriamente) de participantes expuestos a "A". De forma similar, se estima la frecuencia esperada de "B" en relación con "C" midiendo la ocurrencia de "B" después de la exposición a "C" en otra muestra aleatoria de los participantes en el estudio. Esa segunda muestra representativa está conformada por los restantes sujetos, ese subgrupo que también se conforma por azar, al ser "no escogidos" aleatoriamente para exponerse a "A". Luego se comparan esos dos estimadores no sesgados (medidos en muestras aleatorias, no sesgadas), para estimar la dirección y la magnitud de la supuesta asociación entre "A" y "B".

En contraste, en los estudios observacionales, la estimación del efecto no se hace en una muestra probabilística de pacientes expuestos tomados de todo el grupo de susceptibles incluidos en el estudio, sino que se documenta la presencia de la exposición, la cual ocurre sin intervención del investigador. Dicha exposición puede ser un atributo (edad, sexo, grupo sanguíneo) o un *factor externo* (fumar, sedentarismo, tomar analgésicos), que, simplemente, se halla ausente o presente en un grado variable en los sujetos de la población fuente. Por tanto, comparado con el experimento, el estudio observacional analítico no ofrece una estimación intrínsecamente válida (libre de sesgo de asignación) del efecto. Particularmente, cuando hay diferencias sistemáticas en la distribución de factores que puedan afectar el desenlace entre quienes se hallan expuestos o no al factor de interés, se puede producir un *sesgo de confusión*.

En síntesis, dado que en los estudios observacionales la asignación a la exposición no es experimental aleatoria, el investigador debe poner un particular empeño en controlar las fuentes potenciales de confusión y de otros sesgos, así como en el ensamblaje, la evaluación y el análisis de los resultados observados en la muestra estudiada. Siempre se debe medir y controlar el efecto de otros factores diferentes de la variable de exposición que podrían confundir o modificar las asociaciones a los desenlaces. Eso se puede lograr utilizando estrategias de control en el diseño y el ensamblaje del grupo de estudio (restricción, apareamiento) o en el análisis (estratificación, ajustes multivariados).

Dadas las limitaciones prácticas y éticas de los experimentos en seres humanos, los estudios observacionales analíticos juegan un papel de vital importancia para aprender sobre factores de riesgo, causales o pronósticos, e, incluso, para evaluar el desempeño (no la efectividad) de intervenciones.

ASOCIACIÓN Y CAUSALIDAD: FACTOR CAUSAL, FACTOR DE RIESGO Y FACTOR PRONÓSTICO

En la investigación explicativa se busca establecer si la exposición a un factor es ***causa*** de un desenlace. También se busca esclarecer el mecanismo por el cual la exposición al factor ***causa*** el desenlace de interés. A pesar de la intención de generar explicaciones a los eventos observados, en muchas ocasiones la investigación solo se limita a evidenciar la existencia de una asociación entre un factor y un desenlace, sin poder establecer si dicha asociación es causal.

Se define como ***factor de riesgo*** un evento, un fenómeno o un atributo que se asocia a una mayor frecuencia (un mayor "riesgo") de aparición de enfermedad, e, incluso, de muerte por dicha enfermedad, en una persona potencialmente susceptible. Cuando la exposición a un

factor se asocia a una disminución del riesgo, se habla de un ***factor protector***.

En el caso de que consideremos a un sujeto ya enfermo, el evento que se quiere predecir o explicar es el desenlace de la enfermedad: curación, cronicidad, secuelas, muerte, etc. Los factores que se asocian a cambios en la frecuencia (cambios en el "riesgo") de presentación de dichos desenlaces se conocen como ***factores pronósticos***. Hay factores ***de mal pronóstico*** y *de buen pronóstico*, equivalentes a los de riesgo y a los protectores. De hecho, un mismo factor puede ser de riesgo para presentación de la enfermedad (por ejemplo, la hipertensión arterial para el infarto de miocardio), y, al mismo tiempo, ser factor pronóstico: dado que se acaba de padecer un infarto, tener hipotensión es factor de mal pronóstico, mientras que tener la tensión arterial normal o elevada se asocia a una mayor probabilidad de supervivencia.

La **figura 17.1** ilustra un modelo lineal de historia natural de la enfermedad y del papel de factores asociados a su desarrollo. La investigación observacional, tanto descriptiva como analítica, encuentra su nicho más adecuado en la descripción y el análisis de la historia natural de la enfermedad identificando, prediciendo y explicando el papel de factores asociados a la aparición de enfermedad y su tránsito a través de las diferentes etapas de su historia natural, o *curso clínico*, hasta su desenlace final: curación, muerte, cronicidad o secuelas. Por otra parte, las intervenciones en salud buscan modificar favorablemente la historia natural o el curso clínico de la enfermedad, y en la **figura 17.1** se ilustran como diferentes niveles de prevención. El estudio del papel de intervenciones enfocadas en modificar el riesgo de aparición, manifestación, agravamiento y muerte o secuelas por enfermedad, daño o lesión en seres humanos se

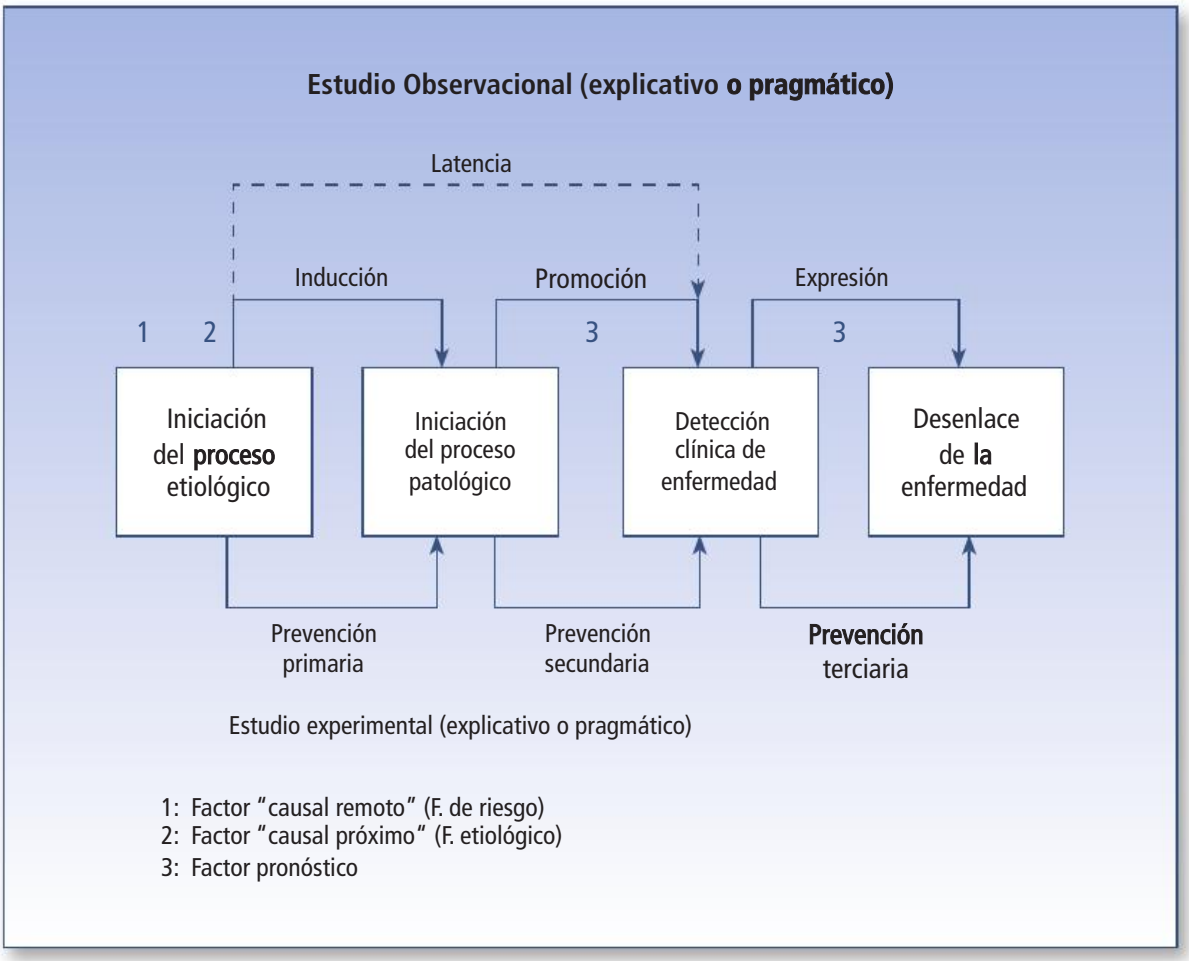


FIGURA 17.1 Papel de la investigación observacional en el estudio de los factores y las etapas de la historia natural, y curso clínico de la enfermedad.

debe hacer, preferentemente, con investigación experimental.

La asociación entre los desenlaces y los factores tanto de riesgo como pronósticos no necesariamente es causal. El factor puede ser, simplemente, un marcador de la presencia de exposiciones, hechos o circunstancias que sean los verdaderamente asociados causalmente al desenlace. Remover un factor de riesgo que no está asociado causalmente puede no reflejarse en un cambio en la probabilidad de presentación del evento. Para determinar si una asociación es de tipo causal, se requiere una serie de criterios que van más allá de la simple asociación, y los cuales son discutidos ampliamente en otras fuentes (Rothman, 1998).

ESTUDIO DE COHORTES

En epidemiología, demografía, ciencias sociales y bioestadística, una cohorte es un conjunto de seres humanos, donde los elementos de dicho conjunto son homogéneos respecto a alguna característica básica.

El estudio de cohortes es particularmente adecuado para evaluar factores de riesgo, factores causales y factores pronósticos, pues, como ya se mencionó, gracias a su diseño, evita las dificultades éticas y logísticas de la investigación experimental del riesgo y causalidad en seres humanos. Fundamentalmente, se trata de ensamblar un grupo muy similar de sujetos (cohorte, del latín ***cohors***, término que se empleaba para designar a grupos muy homogéneos —particularmente, en el ejército romano—) y dividirlos entre quienes se hayan puesto en contacto o no con el factor de riesgo en sospecha. Se hace un seguimiento a los distintos grupos y se registra la frecuencia de presentación del desenlace de interés y sus posibles diferencias. Los resultados muestran cuántas veces más común es el desenlace en el grupo expuesto al compararlo con el grupo no expuesto.

Concepto de cohorte y estudio de cohorte descriptiva

En epidemiología, se suele definir una *cohorte* como un grupo de sujetos que comparten al menos dos atributos: a) son ***susceptibles***; es decir, tienen la posibilidad (están en riesgo) de presentar un desenlace específico (enfermedad, complicación, secuela o muerte), y b) en el momento de ingresar a la cohorte (***tiempo 0***) están libres del desenlace de interés.

Al satisfacer ambos requisitos, el seguimiento de una cohorte permite establecer la frecuencia con la que aparecen los desenlaces de salud en los sujetos susceptibles o en riesgo, a lo largo del tiempo. Esta proporción de eventos de interés que ocurren a lo largo de un periodo se conoce como ***incidencia acumulada***, y es la forma convencional y aceptada de estimar la probabilidad o el riesgo simple de desarrollar un desenlace específico, dado que se tienen ciertas características (***se pertenece a una cohorte específica***).

EJEMPLO

Se ensambla una cohorte de 1000 mujeres adolescentes y adultas jóvenes, quienes reportan tener vida sexual activa al entrar a estudiar a una universidad. Al cabo de un año, se encuentra que 66 de ellas reportaron haber tenido al menos un embarazo no deseado desde cuando fueron admitidas a la universidad. La frecuencia observada de embarazo en universitarias jóvenes que ingresan a dicha institución es del 6,6 % por año. Esa información puede usarse para estimar el riesgo (la probabilidad) de que una estudiante universitaria joven, con vida sexual activa y que ingresa a una universidad similar presente un embarazo no deseado durante el primer año.

En este ejemplo se hizo un estudio de incidencia o riesgo, en una cohorte. El estudio se limita a describir la frecuencia observada del desenlace de interés, y se puede clasificar como un estudio observacional, descriptivo, de cohorte.

Cohorte analítica

En el estudio del ejemplo anterior, no se evaluaron otros factores que podrían modificar la probabilidad de embarazo y estar presentes en algunas de las participantes, o en todas. Se puede conjeturar que el riesgo de embarazo difiere entre quienes usan algún método anticonceptivo y quiénes no. Otros factores también podrían influir: la frecuencia de relaciones sexuales, el método anticonceptivo empleado, el cumplimiento con el método, etc.

EJEMPLO

Supongamos que un investigador se pregunta si las estudiantes que vienen de otras ciudades tienen más riesgo de presentar un embarazo no deseado. Esa conjetura puede provenir de observaciones anecdóticas que sugieran que hay más embarazos en las estudiantes de fuera de la ciudad, o de la creencia de que por vivir fuera de su casa la frecuencia de relaciones sexuales sea más alta, o que por tener una red social diferente (no estar con su familia) su acceso a métodos anticonceptivos sea menor. Al tratarse de un atributo, es obvio que no se puede asignar experimentalmente (no se puede "obligar" a una estudiante local a "venir de fuera"). El investigador puede, entonces, en la cohorte ya ensamblada de 1000 estudiantes, medir la procedencia de las participantes. De esta manera, las puede clasificar y dividir en 2 grupos, de acuerdo con su procedencia: vienen a estudiar desde otra ciudad (exposición de interés) o estudian en su ciudad de residencia habitual (exposición de control, o comparador). El investigador puede medir la frecuencia de embarazos no deseados en ambos grupos y contrastarla.

En la variante introducida en este último ejemplo, después de ensamblar la cohorte (sujetos susceptibles, pero libres del desenlace), el investigador identificó quiénes se hallaban expuestas al factor de interés; es decir, hizo una asignación observacional a la variable de exposición (midió la presencia o la ausencia del factor de interés). En este caso, se trata de un estudio observacional comparativo o analítico, de cohorte.

Estudio observacional analítico, de una cohorte

El estudio anteriormente descrito es observacional analítico, y se hizo en una sola cohorte. El estudio de una cohorte, además de que permite estimar y comparar el riesgo del desenlace en ambos grupos, permite calcular la frecuencia (prevalencia) del factor de riesgo de interés. Este diseño puede resultar poco eficiente cuando la prevalencia del factor de exposición es baja. Si en el ejemplo anterior la prevalencia observada de participantes oriundas de otras ciudades fuera solo del 5 %, querría decir que en el grupo de "expuestas" solo habría 50 mujeres, y la estimación de la incidencia de embarazo en ellas sería muy imprecisa.

Estudio observacional analítico, de múltiples cohortes

El investigador puede hacer otro tipo de estudio de cohorte. Puede ensamblar, por ejemplo, dos cohortes. La primera, integrada por mujeres universitarias jóvenes, que cumplan tres condiciones: a) susceptibilidad (vida sexual activa, que hace factible el evento de embarazo no deseado), b) libres del desenlace (no embarazadas al ingreso a la cohorte) y c) que vengan de otra ciudad. Las participantes se seleccionan con base en su exposición. La comparación se hace con la segunda cohorte, seleccionada con los mismos criterios, excepto que en este caso las universitarias no vendrían de fuera de la ciudad. El observado es un estudio observacional analítico de múltiples cohortes (al menos, dos).

En este caso, no se puede calcular la prevalencia del factor de exposición en la población fuente de la cohorte, la cual, de hecho, queda fijada artificialmente por la relación de tamaño entre las cohortes ensambladas. Si, por ejemplo, ambas tienen el mismo número de sujetos, la "prevalencia" sería del 50 %. Por otra parte, a diferencia del estudio de una cohorte, se puede tener de forma eficiente un tamaño de muestra suficiente en cada grupo que se va a comparar, de modo que los estimativos de asociación sean más precisos.

¿Prospectivo o retrospectivo? El tiempo y el estudio de cohorte

El tiempo es una variable que se relaciona con casi todos los aspectos de los estudios de cohorte. Puede ser, simplemente, un marco de referencia (fecha o época de realización del estudio); o puede ser una variable de exposición o, al menos, formar parte del constructo de la variable de exposición (edad de los sujetos, duración de la exposición); o puede ser el desenlace principal (tiempo hasta la curación, tiempo libre de recaída), o puede ser un factor de confusión o modificador de efecto (efecto de cohorte de nacimiento, interacción entre factor de exposición y duración de otro factor).

Tiempo "causal", dirección temporal

Para evaluar si hay una asociación entre un factor (causal, de riesgo o pronóstico) y un desenlace en salud, se compara la frecuencia o el riesgo (incidencia) en pacientes expuestos y

no expuestos, al cabo de un tiempo suficiente (tiempo ***de incubación***, o ***de latencia***) para que el posible efecto se manifieste. La asociación se evalúa partiendo del nivel de exposición en el momento de ensamblar la cohorte (tiempo 0) y documentando si el desenlace aparece en el futuro. Sin importar el tiempo calendario, la asociación es seguida desde la causa (pasado o presente) hasta el efecto (futuro). Se han utilizado diferentes términos para denominar este concepto: ***longitudinal***, ***prospectivo***, o ***de cohorte***.

La **figura 17.2** esquematiza la disposición de los grupos y la direccionalidad "causal" de la evaluación de la asociación entre exposición y desenlace.

EJEMPLO

En el hipotético estudio de cohorte de 1000 mujeres estudiantes universitarias, el estudio se inició con el comienzo del primer semestre académico de 2014. Las participantes fueron entrevistadas durante los primeros días de clases evaluando su procedencia (factor de exposición de interés), si tenían vida sexual activa, la frecuencia y el número de compañeros sexuales, sus métodos de anticoncepción y la frecuencia de uso, entre otros factores. Durante el año se registró mensualmente si se habían presentado un embarazo y si este era deseado.

En el anterior ejemplo, se puede afirmar que el seguimiento es "prospectivo", pues se va hacia delante en el tiempo, desde la exposición (ingresar a la universidad viniendo de otra ciudad) hacia el desenlace (embarazo no deseado). Debido a otras posibles acepciones del término "prospectivo", su uso puede llevar a confusión. Por ejemplo, este uso del término "prospectivo" no implica que la recolección de datos sea "prospectiva".

EJEMPLO

Los investigadores revisaron los registros del servicio médico universitario del grupo de estudiantes admitidas en 2010. En ellos encontraron que 500 estudiantes al ingreso declaraban tener vida sexual activa y mencionaban la frecuencia aproximada y el uso o no uso de métodos anticonceptivos. Revisando los registros de matrículas, pudieron establecer la procedencia de dichas alumnas. Las historias clínicas, además, contenían información sobre la presencia o no de embarazo durante el año lectivo 2010. En los casos donde hubo registro de embarazo, se hizo contacto telefónico con las estudiantes y se les preguntó si el embarazo había sido no deseado. Con estos datos, se cuantificó la incidencia del embarazo no deseado en las universitarias locales y de fuera de la ciudad en la cohorte de 2010, y se pudo compararla con la cohorte de 2014, que se acababa de estudiar.

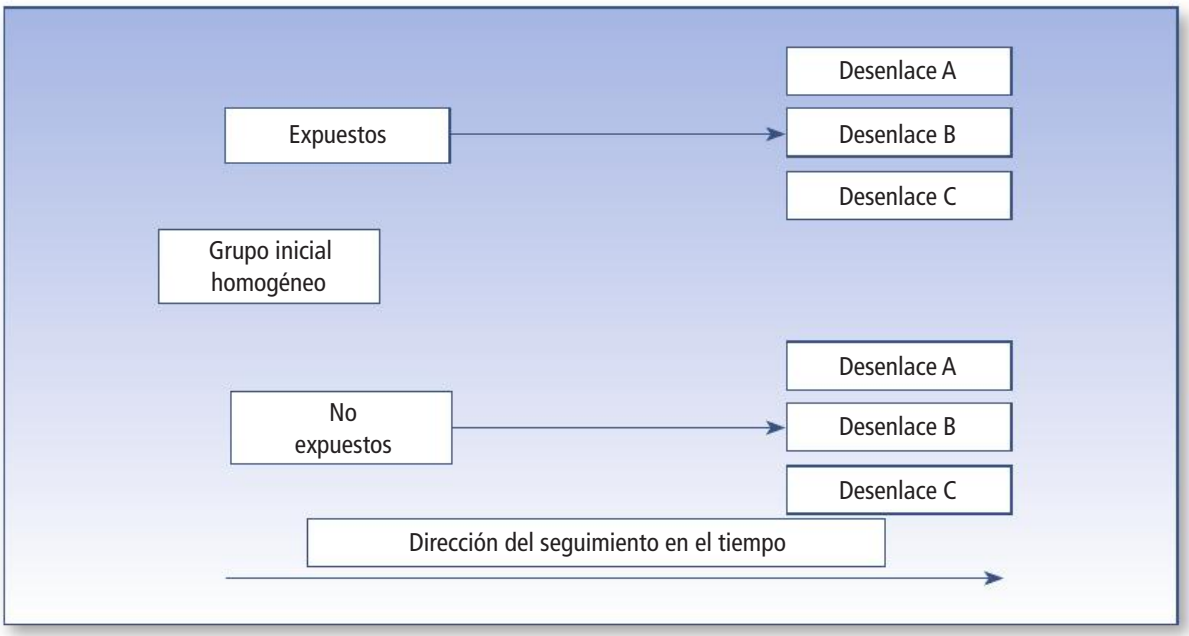


FIGURA 17.2 **Representación esquemática de la arquitectura de un estudio de cohorte analítica, que evalúa la aparición futura** de desenlaces en sujetos en riesgo con niveles diferenciales de exposición.

Tiempo calendario y el ensamblaje de la cohorte

Como se ilustra en el ejemplo anterior, puede perfectamente ocurrir que algunos de los participantes en un estudio, o todos, hayan iniciado la exposición en el pasado o hayan sido seguidos por médicos hasta documentar si, al cabo de un tiempo, presentaron o no el desenlace; y todo eso haber ocurrido antes de la iniciación del estudio y haber quedado registrado en historias clínicas. Cuando el investigador inicia su estudio, la recolección de los datos va a ir desde el presente hacia el pasado (extrayendo de las historias clínicas los eventos pretéritos relacionados con exposición y desenlace), aun cuando el análisis sigue siendo en la dirección de la exposición ("pasado" causal) hacia el desenlace. Desde el punto de vista de captura de la información, la dirección del estudio es "retrospectiva", pero desde el punto de vista de la evaluación de la relación entre exposición y desenlace, el estudio es "prospectivo", lo cual puede resultar confuso.

Para evitar ese tipo de confusión, Alvan Feinstein (1985) propone el uso de los términos ***prospectivo*** y ***retrospectivo*** para referirse al concepto de seguimiento de la cohorte desde su incepción hasta el desenlace (prospectivo) o desde el desenlace hasta la exposición (retrospectivo), independientemente de si los datos son recogidos desde el presente hacia el futuro o si la información fue recogida en el pasado y solo es analizada en el presente. Para referirse a la relación entre la recolección de los datos y el tiempo cronológico, Feinstein usa el término ***prolectivo***, para indicar que los datos se recolectan a partir del momento presente y se "leen" hacia el futuro. En caso de que la información ya exista (es decir, que los eventos estudiados hayan ocurrido en el pasado), usa el término ***retrolectivo***, leyendo la información del presente hacia atrás.

Aunque lógica, la propuesta de Feinstein no ha tenido una acogida generalizada, por lo cual, y para evitar confusión, en este libro se usa el término ***longitudinal***, o, simplemente, *estudio de cohorte*, para indicar que la dirección causal es "prospectiva", o de exposición, hacia el desenlace. Si el ensamble y el seguimiento de la cohorte ocurrieron en el pasado, la recuperación de la información irá del presente hacia el pasado (retrolectiva de Feinstein), y entonces se habla de una ***cohorte histórica***. Si la cohorte se ensambla en el presente y se sigue hacia el futuro (prolectiva de Feinstein), se habla de una *cohorte concurrente*. Puede ocurrir que la cohorte se haya iniciado en el pasado, pero en el momento de empezar la investigación no se haya completado el período de seguimiento y se siga con recolección prolectiva de los datos. En este caso, se trata de una ***cohorte ambidireccional*** **(figura 17.3)**.

En contraste con el concepto anterior, un estudio que identifique a sujetos con y sin el desenlace de interés (casos y no casos, o "controles") y busque asociación al antecedente de exposición o no al factor de riesgo, tanto en los casos como en los controles, tendría una dirección causal "retrospectiva". Seguiría siendo "retrospectivo" independientemente de si la recolección de los datos es "prospectiva", o, según Feinstein, "prolectiva" (como ocurre en estudios de casos y controles incidentes) o "retrospectiva" (retrolectiva). A dichos estudios, que van del desenlace a la exposición, no los llamamos "retrospectivos" en el presente libro: los denominamos, simplemente, ***estudios de casos y controles***.

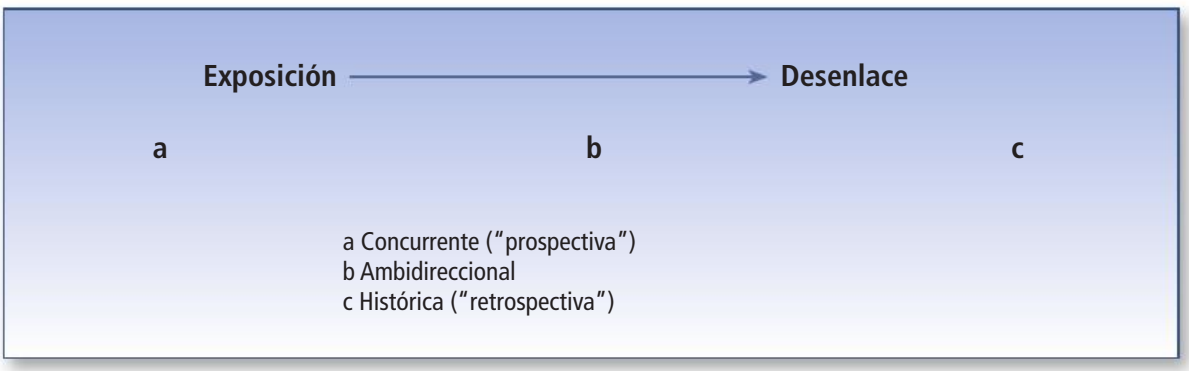


Figura 17.3 Denominación de una cohorte, según la temporalidad de la recolección de la información.

Tiempo de incepción, tiempo de seguimiento, tiempo de incidencia y ensamblaje de la cohorte

Recapitulando, un estudio longitudinal, o de cohorte, se basa en medir la incidencia del evento de interés en un grupo de sujetos susceptibles, pero libres del desenlace al momento de ingresar a la cohorte. El tiempo hasta el evento de interés puede ser el desenlace principal, tal como ocurre en los estudios de supervivencia o de tiempo libre de recaída. Cuando el tiempo hasta el evento es uno de los desenlaces principales, la cohorte debe ensamblarse de tal forma que al ingreso los sujetos estén en un mismo punto de la historia natural o del curso clínico de la condición de interés: de otro modo, no se podría medir el verdadero tiempo transcurrido desde un punto específico de la enfermedad hasta el evento, y las comparaciones de duración de supervivencia no se podrían interpretar adecuadamente.

EJEMPLO

Una investigadora interesada en evaluar si la edad al momento de hacerse el diagnóstico de leucemia linfoide aguda (LLA) en niños menores de 15 años es un factor que afecta el curso clínico de la enfermedad, decide hacer un estudio de cohorte concurrente. Durante enero, febrero y marzo de 2008, recluta a niños que ingresan al servicio de oncología de dos hospitales pediátricos universitarios. De acuerdo con la edad a su ingreso al estudio, los clasifica como menores de 5 años y de 5 a 14 años cumplidos. En total, recluta a 24 pacientes con diagnóstico confirmado de LLA; 10 ingresan en enero, ninguno en febrero y 14 en marzo; los sigue mientras reciben su tratamiento y hasta cuando fallezcan por la leucemia o su tratamiento, hasta cuando presenten recaídas de la enfermedad o hasta cuando permanezcan libres de recaídas hasta por 5 años luego de iniciar el tratamiento. Durante las etapas de inducción y de mantenimiento de la terapia, 5 de los pacientes fallecen, dentro de los 6 primeros meses de seguimiento. Los otros 19 consiguen entrar en remisión de la enfermedad. Uno más recae a los 24 meses del diagnóstico, y fallece. Otro recae a los 37 meses, y los demás están vivos y libres de LLA a los 60 meses (5 años) de seguimiento. El estudio termina el 1 de abril de 2013, 5 años después de que ingresó el último paciente.

En este ejemplo, los sujetos entran a la cohorte en diferentes fechas, pero la cohorte se considera ensamblada no en una fecha determinada, sino en un momento definido del curso clínico de la LLA. Dicho "punto de partida" ocurre cuando, después de confirmado el diagnóstico, cada paciente inicia el tratamiento. Los tiempos de seguimiento (tiempos en riesgo) de cada participante se cuentan a partir de ese momento.

Una cohorte que, sin importar la fecha (tiempo calendario) en la cual ingresen sus participantes, considera que su tiempo 0 (momento del ensamble) es la ocurrencia de un evento significativo y uniforme en la historia natural o el curso clínico de la condición de interés se denomina ***cohorte de origen*** o ***cohorte de incepción*** (anglicismo derivado del término *inception cohort*). El tiempo transcurrido entre el momento 0 (origen o incepción) y la ocurrencia del evento se denomina ***tiempo de incidencia***. Si el evento incidente de interés es la muerte, el tiempo de incidencia corresponde a la duración del tiempo de supervivencia (***survival time***). Si el evento es una recaída o una complicación, el tiempo incidente recibe la denominación apropiada para el evento; por ejemplo, ***tiempo libre de recaída***.

La **tabla 17.1** muestra los intervalos de tiempo calendario entre la iniciación del tratamiento y el fallecimiento o la remisión en los cinco sujetos que murieron dentro de los primeros seis meses de seguimiento y uno de los que sobrevivieron los primeros seis meses.

Si el tiempo hasta el fallecimiento se cuenta desde el inicio del estudio (enero de 2008), en los tres pacientes fallecidos que ingresaron en marzo se habría sobreestimado la duración de la supervivencia en dos meses para cada uno (segmentos marcados en gris). Para hacer la estimación correcta, es necesario ensamblar la cohorte en el momento de iniciar el tratamiento, y no en el mes calendario durante el cual se inicia el estudio **(tabla 17.2)**.

En la **tabla 17.2**, los segmentos coloreados de azul muestran el tiempo, en meses cumplidos, durante los cuales cada sujeto fue seguido. Durante dicho tiempo de observación, cada sujeto está en riesgo de presentar recaída o muerte. Si se suman los periodos resaltados en azul aportados por los cinco participantes fallecidos, se tiene el tiempo total "en riesgo" con el que cada uno contribuyó al estudio. A partir del evento

Tabla 17.1 Fecha (mes/año) de ingreso y fecha (mes/año) de fallecimiento en cinco sujetos del hipotético estudio de cohorte sobre supervivencia libre de recaída en niños tratados para LLA. El paciente 6 está vivo y en remisión en agosto de 2008

Enero 2008	Febrero 2008	Marzo 2008	Abril 2008	Mayo 2008	Junio 2008	Julio 2008	Agosto 2008
Paciente 1			Fallece				
Paciente 2					Fallece		
		Paciente 3	Fallece				
		Paciente 4			Fallece		
		Paciente 5				Fallece	
		Paciente 6					Vivo y en remisión

(en rojo), el sujeto cesa de estar en riesgo y el tiempo posterior ya no es tenido en cuenta. El tiempo total en riesgo de los 5 fallecidos es 16 meses, y, en promedio, cada uno aportó 16 / 5 = 3,2 meses. Este tiempo en riesgo hasta la ocurrencia del evento, para estos 5 sujetos, es el mismo tiempo incidente. Para el sexto sujeto, quien a los 6 meses de seguimiento está vivo, ocurren dos cosas que cabe resaltar: a) aporta, hasta ese momento, 6 meses en riesgo al tiempo total en riesgo de la cohorte estudiada; y b) el tiempo hasta el evento, o tiempo incidente, no ha podido ser determinado, pues el evento no ha ocurrido durante el periodo de observación; solo sabemos que es superior a 6 meses.

Proporción de incidencia y tasa de incidencia

Cuando el tiempo incidente hasta un evento no puede ser determinado en todos los sujetos, bien sea porque se completó el tiempo de seguimiento y no se presentó aún el evento, o bien, porque el sujeto dejó de hallarse disponible para observación (abandona el estudio, fallece por una causa no relacionada, etc.), la información de la cohorte debe resumirse de otra manera. Una forma consiste en calcular la proporción de sujetos en riesgo que presentaron el desenlace en algún momento del intervalo observado. En el numerador se cuentan los sujetos con el desenlace, y en el denominador, todos los sujetos en riesgo al comienzo del periodo, y cuyo destino se conoce al final del intervalo observado. Esta medida de resumen es la ***proporción de incidencia***, o ***incidencia acumulada***, a la cual nos hemos referido previamente en este capítulo. La proporción solo se puede calcular en los sujetos en quienes hay información de cómo están al final del periodo de observación.

Si el destino de algunos de los miembros de la cohorte no es conocido al final de intervalo de observación, en vez de excluir a quienes tienen seguimientos inferiores al periodo estipulado (lo que disminuye el tamaño de muestra, y

Tabla 17.2 Cohorte de origen: mes de ingreso y mes de fallecimiento en cinco sujetos del hipotético estudio de cohorte sobre supervivencia libre de recaída en niños tratados para LLA. El paciente 6 está vivo y en remisión al cabo de 6 meses

Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6
Paciente 1			Fallece		
Paciente 2					Fallece
Paciente 3	Fallece				
Paciente 4			Fallece		
Paciente 5				Fallece	
Paciente 6					Vivo y en remisión

podría, además, sesgar la estimación de la frecuencia), se puede utilizar otra forma de resumir la información: calcular la tasa de incidencia, la cual no es sino el cociente o la razón entre el número de eventos (desenlaces de interés) y la cantidad de tiempo-persona en riesgo acumulada durante el intervalo observado.

EJEMPLO

En la **figura 17.5** se muestra el seguimiento hasta los 6 meses de 6 sujetos. El tiempo promedio hasta el evento en los 6 no puede ser calculado, porque el sexto sujeto, al final del periodo (6 meses) no ha presentado el evento de interés (muerte). Se puede calcular la proporción de incidencia: 5 sujetos fallecidos / 6 sujetos en riesgo, observados hasta fallecer o hasta completar 6 meses. También se puede calcular la tasa de incidencia: 5 fallecimientos / 22 meses-persona en riesgo. Cabe resaltar que la proporción de incidencia no tiene unidades, mientras que la tasa de incidencia, sí.

En el estudio hipotético en 24 niños con leucemia descrito anteriormente, de los 19 sujetos que no fallecen a lo largo de los 6 primeros meses y logran entrar en remisión de la LLA, 2 presentan recaídas durante el tiempo de seguimiento y 17 completan el seguimiento sin evidencia de actividad de la enfermedad. Para los 2 que recaen, los tiempos incidentes son 24 meses (el sujeto que fallece luego de haber entrado en remisión y presentar recaída) y 37 meses. En este caso, cada uno de los 17 participantes sin evidencia de recaída contribuye con 60 meses en riesgo, como "aporte" al acumulado de meses/persona observados. La sumatoria de todos los tiempos en riesgo aportados por la cohorte es: 16 meses (los 5 niños fallecidos antes de 6 meses) + 24 meses (el otro fallecido) + 37 meses (el que presentó recaída) + 1020 meses (60 meses × 17 sujetos que completaron el estudio vivos y libres de recaída) = 1097 meses/persona. La tasa de mortalidad (tasa de incidencia de muerte) durante los primeros 5 años después del diagnóstico, estimada a través del mencionado estudio, es de 6 fallecimientos / 1097 meses/persona = 5,47 muertes por 1000 personas/mes; o dividiendo 1097 entre 12 para convertir meses en años, serían 6 muertes / 91,42 años/persona = 6,65 muertes / 100 personas/año en riesgo.

EJEMPLO

Supongamos que de los 17 sujetos vivos y sin recaída, uno fallece en un accidente de tránsito a los 50 meses de haber iniciado el tratamiento. Es imposible saber si, en caso de haber sobrevivido al accidente, habría presentado recaída antes de los 60 meses. Esta observación aporta información sobre tiempo en riesgo y desenlace durante los primeros 50 meses postiniciación del tratamiento, y a partir de su fallecimiento, por una causa no relacionada (mortalidad competitiva), deja de aportar información sobre el riesgo.

La información de este sujeto queda censurada a partir de ese momento. Algo similar ocurre con el paciente que presentó recaída a los 37 meses, aun cuando haya sobrevivido. Al presentar su primera recaída, deja de estar en riesgo para presentar una primera recaída letal o no letal, por lo cual a partir de ese momento deja de aportar tiempo en riesgo; o, dicho de otra forma, ***deja de pertenecer a la cohorte.*** Aun cuando el mencionado paciente siga siendo observado en el estudio, la información que proviene de él a partir del mes 37 de seguimiento se halla censurada, y no aporta al numerador ni al denominador de la tasa de incidencia.

Sesgo de detección temprana (*lead-time bias*)

En los casos en los que se ensambla la cohorte en el momento del diagnóstico, puede generarse una distorsión sistemática del tiempo hasta el evento de interés (sesgo), cuando en algunos de los sujetos se hace una detección más temprana de la enfermedad; por ejemplo, durante el periodo preclínico del padecimiento. Si la probabilidad de recibir un diagnóstico temprano se asocia a uno de los factores cuya exposición está siendo estudiada (por ejemplo, la edad del paciente), se puede generar un tipo de sesgo conocido como ***sesgo del adelanto del tiempo del diagnóstico***, o ***sesgo de detección temprana*** (***lead time bias***). Parecería como si la supervivencia de los sujetos detectados más tempranamente fuera más prolongada, cuando, en realidad, el tiempo de vida desde el inicio del periodo patogénico puede ser igual **(figura 17.4)**. Esto ocurre en el caso de que la detección más temprana, y, por ende, la iniciación más temprana del tratamiento, en realidad, no afecta el curso clínico de la enfermedad.

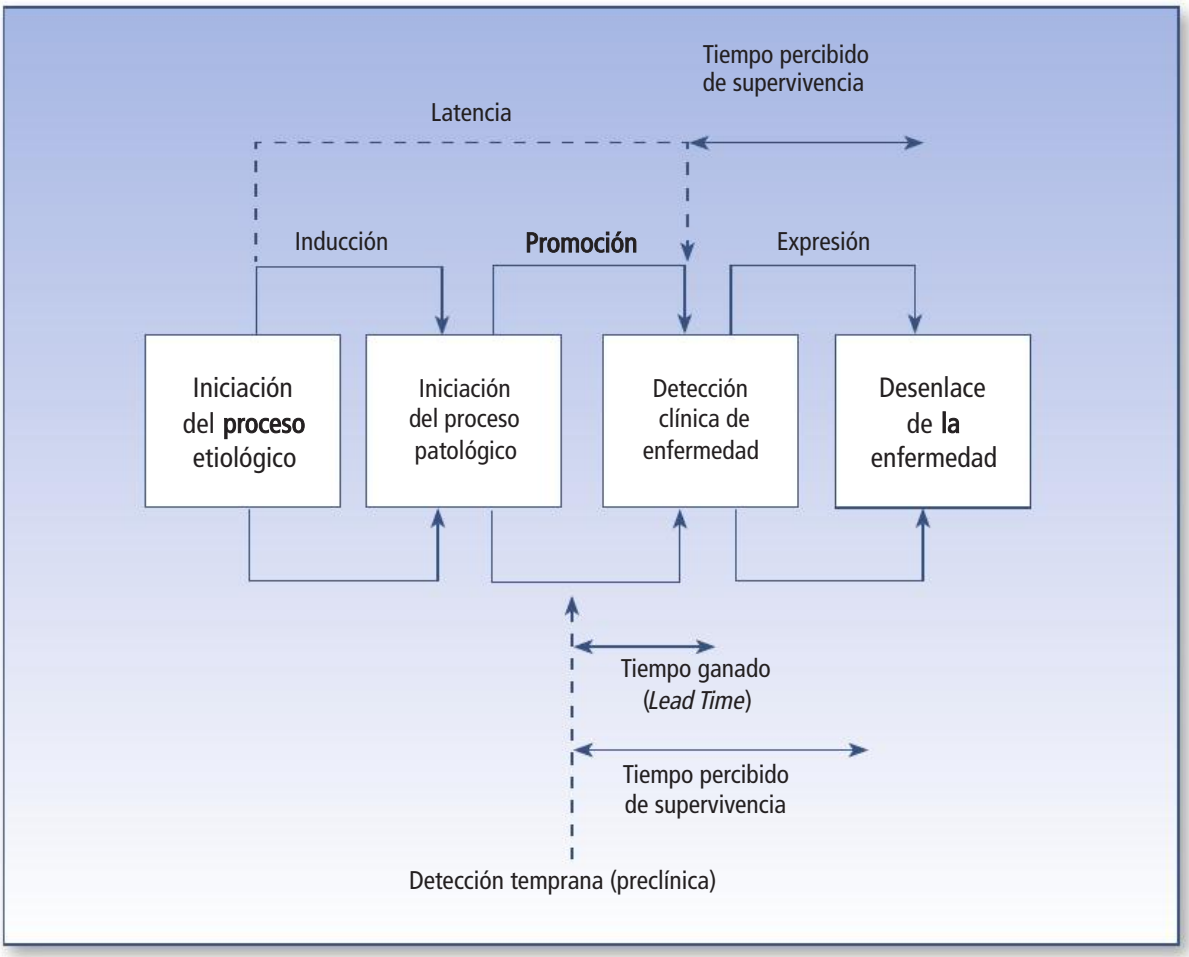


FIGURA 17.4 **Sesgo de detección temprana. La duración real desde el inicio del periodo patogénico no es distinta en los pacientes** detectados durante el periodo preclínico, comparada con la de los detectados durante el periodo clínico.

Etapas del diseño y la conducción de un estudio analítico de cohorte

El diseño y la realización de un estudio de cohorte tienen una serie de componentes y etapas, tal como se ilustra en la **figura 17.5.**

Ensamblaje de la cohorte

El proceso de ensamblaje de la cohorte que será evaluada en un estudio longitudinal analítico se compone de tres etapas: 1) selección, 2) muestreo y 3) asignación.

Selección

Seleccionar consiste en definir un conjunto (***población***) de elementos que tienen las características y las condiciones que se quiere estudiar; o, fraseado alternativamente, ***especifica la población*** *sobre la que el estudio quiere aprender algo*. Dicha población estudiada puede ser un conjunto concreto, definido en tiempo, lugar y características individuales de los sujetos que definen su pertenencia al conjunto. Tal tipo de conjunto real y definido se conoce como ***población blanco***.

EJEMPLO

En el estudio hipotético de riesgo de embarazo no deseado, la fuente de participantes es la consulta de planificación familiar. No todas las estudiantes que tienen vida sexual activa acuden a la consulta. De hecho, quienes van son sistemáticamente más cuidadosas con sus prácticas de anticoncepción, y su riesgo de presentar embarazo no deseado es sistemáticamente menor. Su uso de prácticas anticonceptivas puede ser tan adecuado que haga desaparecer cualquier diferencia real de riesgo que haya entre las estudiantes locales y las de fuera de la ciudad.

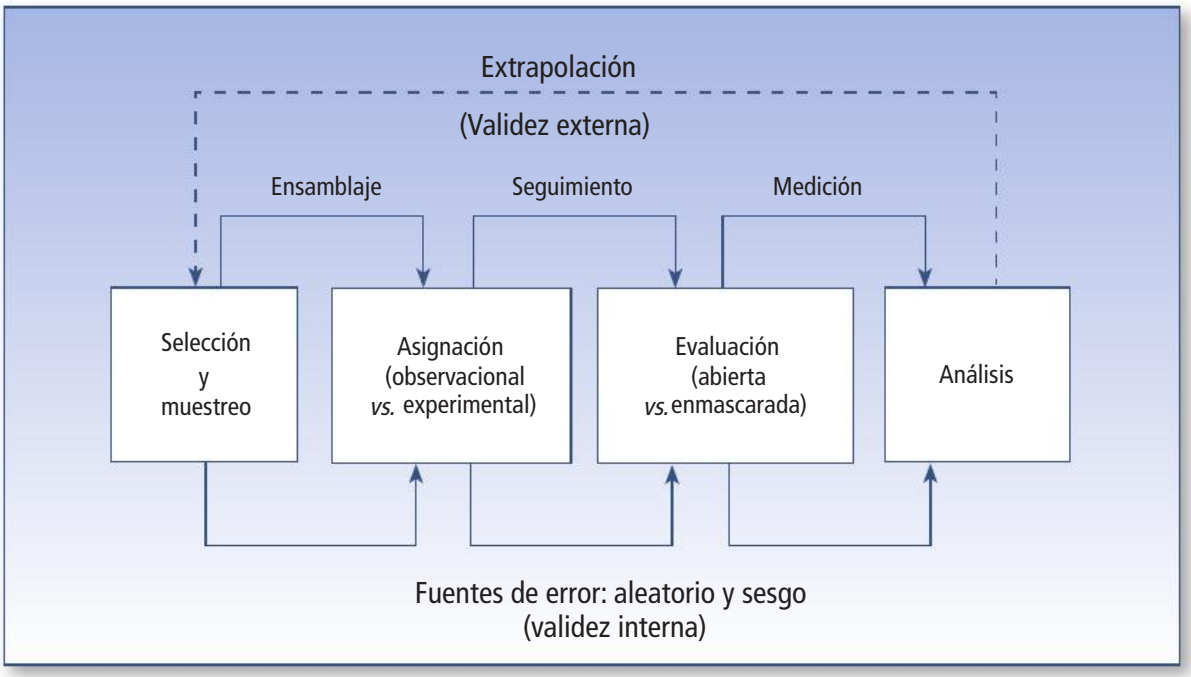


Figura 17.5 Etapas y componentes del diseño, la conducción, el análisis y la interpretación de un estudio de cohorte analítica.

Este ejemplo ilustra una población blanco, definida en términos de tiempo (fechas de ingreso a la empresa), lugar (nombre de la empresa y localización de su sede) y persona (características de las personas: trabajadores sin el desenlace de interés, con riesgo de desarrollarlo). Generalmente, la definición de la población blanco se refina hasta tener criterios detallados de inclusión y exclusión (***población de estudio***, o elegible) y accesibilidad (***población accesible***).

El investigador puede estar interesado no tanto en el riesgo concreto de los trabajadores de una determinada empresa, sino en el concepto general de riesgo de sobrepeso y obesidad cuando el empleo o el oficio de un adulto son sedentarios. En ese caso, puede definir su población de interés en términos abstractos especificando las características relevantes de los sujetos, sin limitarlos a un grupo humano concreto (no definidos en tiempo y lugar, sino solo en persona). En este caso, la población de interés corresponde al concepto de ***población de referencia*** (ver discusión de esos conceptos en el capítulo sobre población y muestra).

Independientemente de si la intención del estudio es estimar un parámetro o un efecto en una población concreta (blanco) o evaluar un atributo o una asociación en una situación abstracta o general (***población de referencia***), el investigador va a necesitar una fuente de sujetos para llevar a cabo su estudio (***población fuente***).

Muestreo

Esa población fuente puede ser la población blanco de interés (o una fracción de ella que sea accesible) o puede concebirse como un modelo o una población de prueba, que represente conceptualmente (no estadísticamente) a la población de referencia, para poder evaluar una hipótesis. De todas formas, el siguiente paso es ensamblar una cohorte con todos los miembros, o una fracción de los sujetos integrantes de la población fuente. Este paso corresponde al proceso de muestreo.

Si la fracción de muestreo es 1 (el 100 % de la población fuente) tanto la precisión (***error aleatorio***, o variación de muestreo) como la validez (***ausencia de sesgo***) de los estimadores obtenidos en el grupo estudiado serán perfectas (para la población fuente; no necesariamente para la población blanco o de referencia originalmente identificada). En muchos casos no es factible, o bien, es ineficiente, incluir a toda la población fuente, y la fracción de muestreo será < 1. La cantidad de sujetos incluidos y la forma de muestreo influyen en la precisión de los estimadores y en su validez.

Para asegurar que los sujetos sean una representación estadística no sesgada de la distribución de los atributos y las características de la población fuente, el proceso de muestreo debe ser aleatorio. Sin embargo, en muchos casos no se requiere representatividad de la población fuente, sino, simplemente, que la muestra de sujetos incluya el rango de las características de interés. En esos casos, basta con hacer un muestreo no probabilístico, muchas veces denominado ***por conveniencia.*** Los muestreos no probabilísticos pueden ser sistemáticos: secuencia de llegada de pacientes, reclutamiento continuo de un grupo o un estrato de sujetos hasta alcanzar un tamaño de muestra específico, etc.

Eventualmente, la cohorte quedará compuesta por un grupo de sujetos susceptibles, pero libres del desenlace, y tan homogéneos respecto a factores demográficos, sociales, económicos y otros condicionantes de la salud como sea necesario, para contestar la pregunta de investigación de interés.

Asignación

En los sujetos seleccionados, se mide la presencia o la ausencia y la intensidad de la exposición al factor que se quiere investigar. Esto divide la cohorte en al menos dos grupos: los pacientes expuestos de forma estable y definida al factor de interés y los que no están expuestos al factor en estudio, ni se prevé razonablemente que lo vayan a estar durante el seguimiento.

En el caso de los estudios de múltiples cohortes, la selección y la asignación se fusionan en una sola etapa, dado que los miembros de cada cohorte son seleccionados de acuerdo con su nivel de exposición al factor de interés.

Asignación (selección) de la población expuesta, en estudios de dos o más cohortes concurrentes

El grupo de la población expuesta puede provenir de distintas fuentes candidatas. Su escogencia dependerá del grado de exposición, de la disponibilidad de información acerca de la exposición, del grado de representatividad de la población de interés para el estudio y de la posibilidad de detectar el evento de interés. Debe ser una población susceptible de seguirse y que brinde la posibilidad de obtener toda la información que el estudio requiera.

Si se evalúa una exposición rara, se puede escoger como población fuente de la cohorte expuesta a un grupo poblacional donde el factor de riesgo esté presente y sea común.

EJEMPLO

Para la evaluación de los efectos de la exposición a cierta sustancia potencialmente tóxica que no es común, se puede seleccionar al grupo de trabajadores de una fábrica que emplee la sustancia, o bien, a la población que vive cerca de un agente ambiental reconocido como un factor de riesgo (cerca de un sitio de almacenamiento o un botadero de basura) o a un grupo de personas con hábitos alimenticios comunes o con estilos de vida similares.

Con el estudio de estas poblaciones, el diseño de cohortes podrá identificar ciertos factores de riesgo que afectan el estado de salud de los individuos; pero, principalmente, el diseño permitirá, al comparar a los grupos expuestos con los no expuestos, detectar los eventos resultantes de la exposición estudiada que, a menudo, podrían pasar inadvertidos en la población general. Aunque el estudio de cohortes, en general, no sirve para evaluar enfermedades raras, por el gran número de pacientes que sería necesario evaluar, a veces, durante tiempos prolongados de seguimiento, sí resulta apropiado para la evaluación de causalidad en las enfermedades con frecuencias no tan bajas.

Asignación (selección) de la población no expuesta, en estudios de dos o más cohortes concurrentes

Se debe escoger una población de sujetos ***no*** *expuestos* **al factor de interés, y que tengan las** mismas características que el grupo expuesto; sobre todo, que tengan una distribución similar de los potenciales factores de confusión identificados por los investigadores. La disponibilidad de la información en el grupo control debe ser comparable a la de la cohorte expuesta. Cuando no se puede definir un grupo de control que satisfaga dichas condiciones, una alternativa viable es seleccionar distintas fuentes de sujetos de control (distintas cohortes de control). Si al comparar la incidencia en la cohorte expuesta con las obtenidas con diferentes grupos no expuestos la dirección de la asociación (por ejemplo, la dirección del riesgo relativo) es consistente,

el investigador puede tener más confianza en que la asociación es real, y no producto de una confusión originada en las características de un grupo de control específico.

En algunos casos, la cohorte puede dividirse para el análisis según los grados de exposición al factor de estudio, si el tamaño de la muestra lo permite. En este caso, se configuran múltiples cohortes, definidas por grados diferentes de exposición y que permiten evaluar relaciones entre dosis de exposición y respuesta (***riesgo del evento*).**

Fuentes de error sistemático en el ensamblaje de la cohorte

Los errores que afecten el ensamblaje de una cohorte pueden afectar la validez interna del estudio o comprometer solo la validez externa de los resultados.

Sesgos en la selección

En primer lugar, cuando se define la población fuente, las distribuciones de las características de este subconjunto pueden diferir sistemáticamente de las distribuciones en la población a la que se quiere hacer inferencia. Estas distorsiones pueden introducirse al especificar los criterios de inclusión y exclusión, o cuando la población accesible es solo un subconjunto específico de la población elegible.

EJEMPLO

En el estudio hipotético de riesgo de embarazo no deseado, la fuente de participantes es la consulta de planificación familiar. No todas las estudiantes que tienen vida sexual activa acuden a la consulta. De hecho, quienes van son sistemáticamente más cuidadosas con sus prácticas de anticoncepción, y su riesgo de presentar embarazo no deseado es sistemáticamente menor. Su uso de prácticas anticonceptivas puede ser tan adecuado que haga desaparecer cualquier diferencia real de riesgo que haya entre las estudiantes locales y las de fuera de la ciudad.

En este ejemplo se ilustra un sistema sesgado de especificación de la población fuente (un sesgo de selección) que causa dos efectos: a) la población fuente es una subrepresentación sistemática del riesgo de embarazo no deseado en la población real de interés —población blanco—, lo cual hace que la extrapolación de los hallazgos a estudiantes que pertenecen a la población blanco (vida sexual activa y riesgo de embarazo no deseado) sea apropiada solo para el subconjunto de las que asisten a una consulta de anticoncepción; y b) debido a que el factor de riesgo más importante para embarazo no planeado es el uso o no uso de métodos anticonceptivos, al hacer que todas las mujeres, independientemente de su procedencia (y de la influencia que el lugar de origen tenga sobre el uso de métodos anticonceptivos), tengan alto nivel de uso de métodos anticonceptivos, atenúa o elimina la asociación entre procedencia y riesgo de embarazo no deseado.

Lo descrito ocurre porque el factor de riesgo de interés (la procedencia) tiene una asociación causal al desenlace, y que no es directa, sino mediada, en gran parte, por el patrón de uso de anticonceptivos (del cual, a su vez, se espera que sea diferente según procedencia). Al seleccionar como población fuente a las pacientes de una consulta de anticoncepción, se elimina este uso diferencial de anticonceptivos según la procedencia de la universitaria, lo cual produce un tipo especial de sesgo que puede ocurrir en la selección o en la asignación, llamado ***sobreapareamiento*.** Su efecto es desviar sistemáticamente los estimativos de asociación hacia la no diferencia entre los diferentes niveles de exposición.

Sesgos en el muestreo

La distorsión sistemática puede aparecer en el proceso de muestreo. En muchas ocasiones, es imposible o impráctico tomar una muestra probabilística de la población fuente. Ello puede ocasionar sobrerrepresentación o subrepresentación de factores de riesgo y pronósticos diferentes del de interés para el estudio, lo que puede distorsionar la verdadera asociación (validez interna) o hacer que la extrapolación a la población de estudio sea inapropiada (validez externa). A diferencia de los ejemplos de sesgos en la selección (los cuales son sesgos en la especificación de la población fuente), la dirección en la que operan los muestreos sesgados no puede ser predicha: hay que evaluarla en cada caso, y, si es posible, prevenir su ocurrencia mediante un proceso de muestreo que represente de forma adecuada a la población fuente.

El investigador, a través del muestreo, invita a los sujetos a participar. Los sujetos, obviamente, son libres de aceptar o no. Hay evidencia empírica de que en una población fuente determinada, los sujetos que aceptan participar en los estudios de salud pertenecen a subpoblaciones sistemáticamente diferentes de las poblaciones de los que no aceptan, respecto a variables tales como hábitos de vida saludables, patrones de uso de servicios de salud y cumplimiento de recomendaciones médicas. La aceptación diferencial (***autoselección** de los participantes*) suele comprometer la validez externa, y, en ocasiones, también la validez interna.

Debido al potencial de sesgo de muestreo introducido por el fenómeno descrito, es aconsejable medir y comparar las características generales de quienes aceptan o no participar.

Sesgos en la asignación

Los sesgos que ocurran en la asignación pueden afectar de forma muy significativa la estimación de las asociaciones estudiadas. Como se mencionó en la introducción, el hecho de que la asignación sea observacional hace que clasificar correctamente la exposición sea uno de los pasos más críticos y vulnerables de toda la planeación y la conducción del estudio de cohorte analítica. Hay dos tipos de error sistemático que afectan la asignación: a) el error en la clasificación y b) el desequilibrio en la distribución de potenciales factores de confusión entre los grupos comparados.

Mala clasificación

Los errores al clasificar la asignación caen en dos categorías: a) ***mala clasificación diferencial*** y b) ***mala clasificación no diferencial.***

Para entender el problema del efecto de la mala clasificación, se debe considerar el efecto de la capacidad discriminativa del sistema de clasificación empleado sobre la asignación de los sujetos. Cuando el método o los métodos empleados tienen una capacidad discriminativa perfecta, aunque se usen métodos diferenciales, todos los sujetos quedan correctamente clasificados, y no hay sesgo en la asignación. El problema es que en la realidad casi ningún sistema de medición puede ser considerado perfecto, y todos tienen algún grado de mala clasificación, bien sea porque su sensibilidad no es perfecta, porque su especificidad no es del 100 %, o porque ambos índices son < 100 % (para una discusión más específica y completa de los conceptos de sensibilidad, especificidad y capacidad discriminativa de una prueba o de un sistema de medición, ver el **capítulo 10** "Uso de pruebas diagnósticas en medicina clínica").

La mala clasificación diferencial ocurre cuando la medición del factor de interés (en este caso, el factor de exposición; pero podría también presentarse en la clasificación del desenlace, como se verá más adelante), se hace usando métodos u observadores diferentes para cada tipo de exposición. Como ya se mencionó, en el mundo real los métodos o los observadores van a producir algún grado de mala clasificación, y al tener capacidades discriminativas diferentes se configura el llamado ***sesgo de mala clasificación diferencial***, que va a clasificar como pacientes expuestos a un subgrupo de personas no expuestas, o viceversa.

EJEMPLO

Un estudio de cohorte busca evaluar la asociación entre el hábito de fumar y la presencia de cáncer de vejiga. Se quiere ensamblar una cohorte de personas sin evidencia clínica de cáncer de vejiga, clasificarlos según sus hábitos de consumo de cigarrillo al comienzo del estudio en "fumadores" y "no fumadores" y seguirlos durante 30 años, para evaluar en ambos grupos la incidencia del cáncer de vejiga. Los investigadores utilizan un algoritmo para clasificar a los sujetos como fumadores o no fumadores. Inicialmente se le pregunta al sujeto si actualmente fuma. Si contesta que sí, es clasificado como "fumador", sin tener en cuenta ninguna otra información. Si contesta que no, se inicia un proceso de validación de que, en verdad, no sea fumador. Para eso se le pregunta si fue fumador y actualmente dejó de fumar. Si contesta que fue fumador, se le clasifica como "fumador", sin tener en cuenta si solo fue un fumador ocasional o si fue un fumador habitual, ni considerar cuánto tiempo lleva sin fumar. Si contesta que no fue un fumador que dejó el hábito, se le pregunta si en algún momento probó el tabaco o ha sido fumador ocasional. Solo si contesta nuevamente que no, es clasificado como "no fumador".

En este ejemplo, la definición de la categoría fumador es muy sensible (no se le escapa ningún verdadero fumador), pero tiene muchos falsos positivos (poco específica). La definición de "no fumador" es, por el contrario, muy específica,

pero poco sensible. Este uso de criterios diferentes para definir a fumador y a no fumador (***clasificación diferencial***) produce una mala clasificación (asignación al grupo incorrecto) sistemáticamente distinta en los grupos de exposición. En este ejemplo en particular, muchas personas que en realidad solo han tenido una exposición mínima y no significativa al consumo de cigarrillo son clasificadas como fumadoras y agrupadas junto con verdaderos fumadores de exposición significativa. Esto "diluye" el efecto del cigarrillo en el riesgo de aparición del cáncer de vejiga y lleva, posiblemente, a subestimar la verdadera magnitud de la asociación (***sesgo hacia la no diferencia***).

Aun cuando se usen los mismos observadores y métodos para definir la exposición presente o ausente en cada sujeto, si el sistema de clasificación es imperfecto (como casi siempre ocurre), se va a producir un sesgo en la estimación del efecto, fenómeno denominado ***sesgo por mala clasificación no diferencial***; en este caso, de la exposición.

EJEMPLO

En el hipotético estudio anterior, se decide clasificar a los sujetos con una sola pregunta, incluida en un cuestionario autoadministrado sin intervención de múltiples observadores, para eliminar cualquier riesgo de mala clasificación diferencial. La pregunta es: *¿Considera usted que ha sido fumador alguna vez en su vida?* La habilidad discriminativa de esta pregunta fue evaluada en un estudio donde se comparó la respuesta que los sujetos daban con una entrevista en profundidad sobre el contacto de cada sujeto con el cigarrillo, más una revisión de sus antecedentes médicos en las historias clínicas y una encuesta a su mejor amigo, preguntando si el sujeto fuma.

Al evaluar el uso de la pregunta simple, se encontró que su sensibilidad era del 100 %: la totalidad de los sujetos clasificados como fumadores a través de la entrevista, la revisión de la historia clínica y el interrogatorio a su mejor amigo contestaron afirmativamente. Por otra parte, la especificidad fue apenas del 50 %. La mitad de quienes la prueba confirmatoria clasificó como sin exposición significativa al cigarrillo, contestaron afirmativamente y fueron considerados falsos positivos.

A pesar de su baja especificidad, los investigadores concluyeron que no se produciría sesgo de asignación, pues aunque habría error en la clasificación, este afectaría simétricamente a los verdaderos fumadores y a los verdaderos no fumadores.

Al clasificar a 2000 participantes usando la pregunta, asignaron a 1500 sujetos a la exposición "fumador", y a 500 más, a la exposición "no fumador". El financiador de la investigación solicitó que se hiciera una validación de la exposición utilizando un sistema de entrevista estructurada, como el que se empleó en el estudio de evaluación de la habilidad discriminativa de la pregunta. Se halló que, en realidad, había 500 no fumadores que habían sido erróneamente clasificados como fumadores.

Este ejemplo ilustra un efecto sobre la composición de los grupos de asignación, producido por una prueba de clasificación imperfecta, aun en el caso de que se use exactamente la misma prueba en pacientes expuestos y no expuestos. Los falsos positivos son la mitad de los verdaderamente no expuestos, que son agrupados con los fumadores, lo cual diluye el efecto de la exposición al cigarrillo, de forma similar a como ocurrió en el ejemplo de mala clasificación diferencial.

El sesgo de mala clasificación no diferencial de la exposición (y también del desenlace) es prácticamente ineludible, debido a que casi ningún sistema de medición está exento de error de clasificación (casi ninguno tiene sensibilidad y especificidad del 100 %, como ya se mencionó). El investigador debe escoger una prueba con la mejor capacidad discriminativa posible, pero, de todas maneras, también anticipar que sus estimativos de efecto podrían estar sesgados. Afortunadamente, en casi todos los casos se puede anticipar que la mala clasificación no diferencial sesga los resultados hacia la atenuación o la desaparición de la asociación (***sesgo hacia la no diferencia***).

Cuando el estudio encuentra una diferencia en las incidencias del desenlace entre pacientes expuestos y no expuestos, el investigador puede afirmar que la diferencia sí existe, pero que su magnitud podría ser mayor que la hallada en el estudio (pese a tener un sesgo operando en su contra, el factor demuestra estar asociado). La mayor dificultad ocurre cuando el estudio no encuentra diferencias, pues la explicación podría ser que, en realidad, no las hay; o que, si las hay, no pueden ser evidenciadas, debido al sesgo hacia la no diferencia producido por la mala clasificación no diferencial.

Desbalance de la distribución de potenciales factores de confusión

Como se mencionó en la introducción, si hay potenciales factores de confusión que es-

tén desbalanceados entre los grupos asignados (observacional o experimentalmente) a las diferentes exposiciones, dichos factores podrían confundir la asociación, y así llevar a distorsionar su dirección y su magnitud verdaderas. La distribución diferencial del potencial factor de confusión ocurre durante el ensamblaje, por la generación de los grupos que se van a comparar a través del proceso de asignación, pero no necesariamente es causada por un error asociado a la asignación. En muchas ocasiones, el potencial factor de confusión, simplemente, covaría con la exposición de interés.

EJEMPLO

En una consulta de mantenimiento de la salud, un médico ha encontrado cómo, a menudo, los pacientes que presentan bronquitis crónica reportan que son tomadores de café. Como el médico no conoce ninguna evidencia que vincule el consumo de café con la bronquitis crónica, decide hacer un estudio de cohorte analítica para evaluar la hipotética asociación entre consumo de café y bronquitis crónica. Cuando ensambla su cohorte, encuentra que los sujetos clasificados como tomadores de café reportan que fuman mucho más a menudo que lo que reportan los no consumidores de café.

Al hacerse, al final del estudio, un análisis estratificando por el factor de confusión (cigarrillo), el médico encuentra que en los fumadores no hay diferencia en la frecuencia de bronquitis crónica entre tomadores y no tomadores de café. Cuando examina el estrato de los no fumadores, el hallazgo es el mismo: no hay relación entre el café y la bronquitis crónica. El investigador concluye que la observación inicial, sugerente de que había más bronquitis crónica en los consumidores de café, era, posiblemente, producto de la confusión con el tabaquismo, pues en esa población los tomadores de café fuman más.

El punto crítico destacado en este ejemplo es que en un estudio observacional, como es el caso de los estudios de cohorte analítica, no se puede asumir que no hay potencial de confusión, por lo cual el investigador debe anticipar la probabilidad de confusión y controlar sus efectos en el diseño (por ejemplo, excluyendo a los fumadores), balanceando su distribución durante la selección o la asignación (apareamiento por parejas o por frecuencias) o controlando su efecto durante el análisis, por estratificación o por ajuste multivariado. Para poder usar cualquiera de esas estrategias, es indispensable que el investigador haya identificado el potencial factor de confusión desde el diseño del estudio y lo mida en los sujetos de la cohorte, bien sea para excluir o para controlar.

Seguimiento

Una vez ensamblada la cohorte y asignados los sujetos, empieza el periodo de seguimiento. La idea es tener información completa, oportuna y adecuada de cómo evolucionan los sujetos, si hay cambios en la exposición, cuánto tiempo en riesgo de presentar el desenlace es observado en cada participante, y, finalmente, si los desenlaces de interés ocurren, y en qué momento.

Como ya se mencionó, la duración esperada del seguimiento de cada participante es parte de la planeación del estudio. Se necesita seguir a los sujetos el tiempo suficiente para permitir que los desenlaces de interés se desarrollen y puedan ser detectados usando los sistemas de medición planeados por el estudio.

Hay varios potenciales puntos de interés durante el seguimiento, que, a su vez, se relacionan con las fuentes y los tipos de error que pueden amenazar la validez del estudio, y deben ser controlados (o, al menos, medidos y reportados): a) permanencia en el grupo de exposición original, b) medición del tiempo en riesgo transcurrido para cada sujeto y c) evaluación: documentación de incidencia o no del desenlace de interés.

Permanencia en el grupo de exposición

En muchos casos, el factor de interés es un atributo que ni el sujeto ni el investigador pueden modificar: sexo, genotipo, raza, etc. En estos casos, es imposible cambiar la exposición durante el seguimiento, y la preocupación por la permanencia del sujeto en el grupo de exposición en el que fue clasificado durante la asignación es irrelevante.

En otras ocasiones, la exposición al factor de interés se refiere a una dosis y una duración preespecificadas que ocurren antes del ingreso del sujeto a la cohorte, y la hipótesis del estudio asume que el efecto atribuible a la exposición ya está determinado al comienzo del estudio (se completó el periodo etiológico), si bien aún no es detectable (periodo patogénico, pero antes de cruzar el umbral de detección o de manifestación clínica).

De acuerdo con lo ilustrado en la **figura 17.1** (al comienzo del capítulo), el supuesto del cual

se parte es que el periodo de inducción ya ha transcurrido y el efecto de la exposición ya está completo. El seguimiento debe cubrir el tiempo necesario para que se complete el periodo de promoción, y así se pueda evaluar si la exposición se asoció o no al desenlace. Este argumento es, cuando menos, discutible. En muchos casos, la exposición continuada durante el seguimiento (el tiempo en riesgo) podría modificar la probabilidad de que se presente el desenlace; y si ese último supuesto es válido, entonces en caso de modificarse la exposición durante el estudio, se podrían alterar o confundir las asociaciones.

EJEMPLO

Se evalúa si alcanzar y mantener niveles normales de hemoglobina glucosilada (HbA1c) durante el primer año después del diagnóstico de diabetes *mellitus* tipo 2 (DM2) es un factor pronóstico que retarda o previene el desarrollo de nefropatía diabética, en pacientes diabéticos sin evidencia de daño renal al momento del diagnóstico de la DM2. Se evalúa a pacientes recién diagnosticados con DM2, y se les mide la presencia de microalbuminuria al comienzo del estudio como marcador de glomerulopatía.

Se incluye en la cohorte a los sujetos que en la evaluación inicial no tienen microalbuminuria. Los pacientes empiezan a recibir manejo estandarizado para su DM2, y cada tres meses durante el primer año de tratamiento, se les miden los niveles de HbA1c. Al cabo de un año, se clasifica a los pacientes en dos grupos: a) *controlados durante el primer año*: los que, al menos durante las tres últimas mediciones, tienen niveles normales de HbA1c; y b) *no bien controlados*: todos los demás. Durante los siguientes diez años, se evalúa cada seis meses la presencia de microalbuminuria en los sujetos de ambos grupos y se les siguen monitorizando cada tres meses los niveles de HbA1c.

En el análisis se comparan las incidencias y los tiempos medios hasta evento de glomerulopatía diabética (presencia de microalbuminuria) en cada grupo definido, según los niveles de HbA1c durante el primer año después del diagnóstico. El análisis muestra que la incidencia acumulada de glomerulopatía diabética en los siguientes diez años no difiere entre sujetos con y sin el adecuado control de la DM2 durante el primer año después del diagnóstico (*RR* 1,6; IC 95 % 0,74-3,5).

Los autores concluyen que alcanzar una HbA1c normal al cabo de un año de tratamiento no protege contra el desarrollo de nefropatía diabética. El manuscrito que describe el estudio es sometido a una revista, para su publicación. Uno de los pares revisores critica el hecho de que los cambios en los niveles de HbA1c ocurridos y documentados durante el seguimiento tras el primer año de tratamiento no hayan sido tenidos en cuenta para evaluar la incidencia de glomerulopatía, y recomienda rechazar el estudio para publicación.

Los autores contraargumentan declarando que su hipótesis tiene que ver con el control inicial de la DM2, independientemente de si no se logra mantener los niveles de hemoglobina glucosilada dentro del rango normal tras el primer año del diagnóstico. Un reanálisis, en el cual los sujetos son clasificados según el tiempo total que cada participante acumula con valores normales de hemoglobina glucosilada durante los 10 años de seguimiento, muestra que cuanto más tiempo tenga un sujeto su hemoglobina glucosilada normal (adecuado control de la diabetes), tanto menor es su riesgo de desarrollar glomerulopatía diabética.

En el ejemplo citado, se ilustran dos conceptos importantes: a) la validez del supuesto de que la exposición de interés ocurre ***antes*** de iniciar el seguimiento, y b) el cambio en los niveles de exposición durante el seguimiento, fenómeno conocido como ***contaminación***: sujetos no expuestos inicialmente se convierten en expuestos, o alternan sus estados de exposición a lo largo del seguimiento. El primer modelo de análisis, en el cual los investigadores analizan a los sujetos en el grupo donde fueron asignados observacionalmente al comienzo el estudio, se asemeja al análisis por "intención de tratar" en los experimentos aleatorios controlados. Mientras en los experimentos este enfoque tiene fuertes razones teóricas y empíricas que lo respaldan, en el caso de los estudios observacionales es discutible, y dependerá de poder demostrar que los cambios en la duración o la intensidad de la exposición durante el seguimiento no influyen en la frecuencia o la severidad del desenlace; un supuesto difícil de sustentar.

Para lidiar de forma eficiente con este problema, generado por la contaminación de la exposición o la fluctuante intensidad de la exposición a lo largo del periodo de latencia, se han propuesto alternativas que toman en cuenta la exposición cambiante: por ejemplo, los ***diseños de panel*** y los ***modelos longitudinales de análisis***.

En el diseño de panel, la cohorte inicial es compuesta por sujetos miembros de una población en riesgo, pero al momento del ensamblaje se incluye a sujetos con y sin el desenlace.

La exposición y el desenlace se miden simultáneamente en los sujetos, en diferentes momentos. Este diseño permite evaluar la frecuencia del desenlace, en contraste con la de la exposición periódica a lo largo del seguimiento.

Los modelos longitudinales de análisis, por su parte, pueden verse como una forma sofisticada de evaluar incidencia, en una aproximación similar al estudio de panel. El modelo longitudinal permite introducir y evaluar el tiempo como parte del desenlace (tiempo hasta el evento), y como exposición (tiempo en riesgo transcurrido en cada nivel de exposición) para cada individuo.

Tiempo en riesgo aportado por cada sujeto

Otro aspecto crítico del seguimiento es verificar el tiempo libre del desenlace, y adecuadamente documentado, que aporta cada participante del estudio. Los conceptos de tiempo de incepción, tiempo en riesgo y tiempo hasta el desenlace ya fueron discutidos en este mismo capítulo.

En este punto, basta con resaltar que se deben hacer esfuerzos para tener el seguimiento más completo posible en cada sujeto. A medida que el estudio avanza, hay una tendencia natural a que disminuya el número de sujetos observados; particularmente, en estudios donde el seguimiento es prolongado (años o décadas). Muchos sujetos presentan el desenlace de interés, y a partir de ese momento pueden dejar de hallarse en riesgo para efectos del estudio. En otras ocasiones, el desenlace puede presentarse muchas veces en un mismo sujeto (por ejemplo, episodios de infección respiratoria aguda) y el desenlace relevante no es la proporción de incidencia, sino la densidad de incidencia (***tasa de incidencia***), que es la razón entre el número total de eventos presentados en un brazo de la cohorte y el tiempo en riesgo en ese mismo brazo. En tales casos, cuando un sujeto experimenta un episodio del desenlace de interés, se lo censura hasta que no se recupere y vuelva a ser susceptible, y a partir de ese momento vuelve a aportar tiempo en riesgo.

Supongamos que un estudio planea hacer un seguimiento por cinco años. Cuando un sujeto fallece al cabo de un año, por causas relacionadas con el desenlace de interés, deja de aportar al tiempo en riesgo. Sin embargo, su seguimiento es completo, y cuando se calcule la proporción de incidencia acumulada hasta el final del seguimiento (por ejemplo, cinco años), aporta al número de sujetos con evento en el numerador, y al número de sujetos en riesgo al comienzo del periodo y con seguimiento completo, que constituyen el denominador.

Por el contrario, cuando un sujeto abandona el estudio a los tres años, sin haber presentado el evento de interés, esa observación queda censurada y no puede ser contabilizada en el denominador cuando se calcule la incidencia acumulada en cinco años. Lo mismo ocurre con un sujeto que sin haber presentado el desenlace de interés fallezca a los dos años por alguna causa no relacionada con la condición bajo estudio (mortalidad competitiva).

Como ya se comentó, estas observaciones censuradas a partir de la pérdida en el seguimiento o del deceso por mortalidad competitiva pueden, de todas maneras, aportar tiempo en riesgo al denominador de una tasa o de una densidad de incidencia contabilizando el tiempo en riesgo transcurrido desde la incepción de la cohorte hasta el evento que origina la censura de la información.

De todas formas, cuando la medida principal de asociación es un riesgo relativo, dichas pérdidas en el seguimiento afectarán primariamente el tamaño de muestra efectivo en cada grupo, y, por tanto, la precisión con la que se pueden estimar tanto las proporciones de incidencia como los estimadores de asociación (precisión del cálculo del riesgo relativo, de la diferencia de riesgos, etc.).

Hay un problema potencial más. Cuando un sujeto abandona un estudio, lo hace por razones diferentes de las que motivan a otro sujeto a seguir en el estudio. Cuando un sujeto de un estudio fallece por mortalidad competitiva (por ejemplo, víctima de un infarto de miocardio), su pronóstico respecto al desenlace de interés (por ejemplo, muerte por cáncer gástrico) podría ser diferente del de quienes no fallecen y completan el seguimiento. En otras palabras, los sujetos con observaciones censuradas, bien sea por abandono, pérdida en el seguimiento o mortalidad competitiva, pueden (y suelen) ser sistemáticamente diferentes de los que completan el estudio hasta presentar el desenlace de interés o completar el seguimiento libre de él. La cohorte "superviviente" es una representación sesgada del verdadero riesgo de la cohorte inicial, al ser despojada sistemáticamente de sujetos con

hábitos, creencias o comorbilidades diferentes de los que se mantienen en el estudio.

Al modificar (por ejemplo, al atenuar) el riesgo global en todos los participantes, esto puede sesgar los estimativos de asociación. El escenario es más complejo si, además, las pérdidas en el seguimiento se distribuyen de forma diferente entre pacientes expuestos y no expuestos, o si las características basales de los desertores (por ejemplo, la edad, la raza, el nivel socioeconómico) difieren entre los grupos de expuestos y no expuestos. Esto puede provocar desbalances en la distribución de potenciales modificadores de efecto, que no se hallaban presentes en el momento de ensamblaje de la cohorte.

Por lo anterior, conviene hacer un análisis de sensibilidad, que evalúe cuánto cambian los resultados cuando se atribuyen desenlaces a los sujetos perdidos en un "mejor" y en un "peor" escenario.

EJEMPLO

Se hizo un estudio para evaluar si el momento del diagnóstico de meningitis bacteriana en niños, a partir de que el paciente presenta fiebre hasta la confirmación por punción lumbar e iniciación de tratamiento adecuado, es un factor pronóstico para sordera neurosensorial posmeningitis. Durante 2 años, en 50 hospitales pediátricos de referencia se recolectaron 1000 casos de meningitis bacteriana, que estaban vivos al momento del egreso hospitalario y cuyos padres reportaron que el niño tenía audición normal antes de la enfermedad. En 500 pacientes, el diagnóstico y el tratamiento fueron tempranos (dentro de los primeros 2 días del inicio de la fiebre) y en los otros 500 fueron tardíos.

Al cabo de 6 meses del egreso, se les evaluó la audición con una medición de potenciales evocados auditivos. Acudieron a la evaluación 450 niños del grupo de tratamiento tardío (el posible factor de mal pronóstico auditivo) y 400 del de tratamiento temprano. El estudio encontró que 90 sujetos del tratamiento tardío tenían sordera (incidencia 90 / 450, o del 20 %), comparado con 40 del otro grupo (40 / 400, o del 10 %). El riesgo relativo (RR) de desarrollar sordera si se tiene tratamiento tardío en el estudio es del 20 % / 10 % = 2 (IC 95 % 1,4-2,8).

El problema con el citado estimativo es que se perdieron del seguimiento 150 sujetos (el 15 % de la cohorte). Un análisis de sensibilidad evaluando las consecuencias del peor escenario, en el cual la probabilidad de ser sordos en los sujetos perdidos del grupo de tratamiento temprano sea del 40 % (40 de los 100 sujetos perdidos en dicho grupo), y solo el 20 % en el grupo de tratamiento tardío (10 de los 50 sujetos perdidos), cambiaría significativamente los estimativos así: incidencia en el tratamiento tardío: 100 / 500 = 20 %; incidencia en el tratamiento temprano: 80 / 500 = 16 %; RR tardío *vs.* RR temprano = 1,25 (IC 95 % 0,95-1,63). El análisis de sensibilidad sugiere que la pérdida diferencial en el seguimiento podría haber provocado que se sobreestimara el efecto del retraso en el tratamiento sobre el riesgo de sordera; y, de hecho, es verosímil que no haya ningún exceso de riesgo (el intervalo de confianza en torno al estimativo del RR incluye el 1).

Cuanto más elevada sea la proporción de sujetos perdidos durante el seguimiento, más amenazada estará la validez del estudio. Se debe tener presente que las pérdidas durante un estudio de cohorte pueden estar relacionadas directamente con la exposición, con los desenlaces de interés, con la comorbilidad y con las causas de mortalidad competitiva, solas o en cualquier combinación. La mejor manera de evitar tal problema es tratar de escoger a una población que pueda ser seguida durante todo el tiempo del estudio y tenga pocas probabilidades de pérdida.

No hay un porcentaje preciso a partir del cual los estudios pierdan validez por las pérdidas de sujetos en el seguimiento. Cuanto más completo sea el seguimiento, mayores serán la calidad y la solidez de los resultados. Esta validez dependerá también de la incidencia de los desenlaces de interés en los grupos: tanto el expuesto como el no expuesto. Cuando la incidencia del evento buscado es baja, los resultados pueden verse afectados por pérdidas relativamente pequeñas de sujetos.

Medición del desenlace

Como en el caso de la asignación observacional, la atribución del desenlace a un paciente es un acto de medición sujeto a los tipos y las fuentes de error que pueden afectar cualquier medida: imprecisión (fluctuación aleatoria) y sesgo, originados en la persona medida, en el instrumento de medición y en el observador. El sesgo de mala clasificación no diferencial se presenta de la misma manera como se describió a propósito de los errores que afectan la asignación observacional.

En el caso del sesgo de mala clasificación diferencial, hay elementos adicionales, depen-

dientes del observador y del sujeto observado, y que amenazan la validez de la atribución de los desenlaces.

EJEMPLO

En un estudio de cohorte analítica se examina la relación entre causa de la muerte y un conjunto de factores de riesgo que incluyen el sedentarismo, el consumo de cigarrillo, el sobrepeso y la obesidad, la hipercolesterolemia y el DM2. Para establecer la causa de muerte, se revisa el certificado de defunción y el informe de la autopsia, cuando esta existe. Si no hay examen post mórtem, el comité de evaluación (que se encarga de definir la causa primaria de muerte para el estudio) revisa la historia clínica del sujeto. A menudo se encuentran casos en los que el certificado de defunción registra como causa de la muerte "paro cardiorrespiratorio", lo cual, por supuesto, no permite detectar la verdadera causa del deceso. En tales casos, el comité infiere la posible causa de muerte a partir de los datos disponibles en la historia clínica y los antecedentes. Por ejemplo, en casos de muerte súbita, el comité suele atribuir la defunción a un evento cardiovascular agudo, si el sujeto tenía antecedentes de hipertensión, sedentarismo, obesidad, tabaquismo o diabetes, los cuales son los mismos factores que están siendo investigados. Si, por el contrario, ninguno de los mencionados factores se halla presente, la clasificación más probable para la muerte será "causa indeterminada".

En este ejemplo, es claro que la asignación de causa de muerte, cuando la información del certificado de defunción es inadecuada y no hay autopsia, va a estar sujeta a un alto grado de imprecisión o incertidumbre. El problema adicional es que el nivel esperado de error o incertidumbre no ocurre aleatoriamente. Como el comité que hace la atribución de la causa de muerte sabe a qué factores de riesgo estaba expuesto el participante, ese conocimiento va a hacer que, en caso de incertidumbre, sistemáticamente se atribuya la muerte a una causa cardiovascular en los sujetos expuestos a los factores de riesgo investigados, lo cual, a su vez, hace que cualquier efecto atribuible a uno de dichos factores de riesgo sea sobrestimado. De hecho, se está incluyendo la exposición al factor de riesgo en la definición de la causa de muerte ("tenía hipertensión, tabaquismo y dislipidemia; por tanto, debió de morir de un infarto").

Si el comité que hace la atribución de la causa de muerte desconoce a qué factores de riesgo de interés se hallaba expuesto el participante, la atribución de la causa de muerte sigue siendo incierta y sujeta al error de clasificación, pero dicho error ya no ocurrirá sistemáticamente en una dirección. La estrategia que se utiliza para evitar o disminuir tal fuente de sesgo por medición diferencial consiste en: a) utilizar los mismos criterios y la misma intensidad de búsqueda del desenlace en todos los sujetos (por ejemplo, siempre solicitar autopsia), y b) hacer que el observador que evalúa la presencia y la intensidad del desenlace de interés desconozca la exposición del participante. Esto es lo que se conoce como ***evaluación ciega o enmascarada*** *del desenlace*.

Análisis

Las estrategias apropiadas para el procesamiento, el análisis, y la presentación de los resultados de estudios analíticos, tanto experimentales como observacionales, se discuten en el capítulo sobre el plan de análisis. Las principales medidas de frecuencia, asociación e impacto que se pueden calcular se discuten, a su vez, en el capítulo sobre medidas de frecuencia y asociación. Adicionalmente, en el capítulo sobre análisis de estudios observacionales se presentan y se discuten en detalle las estrategias para hacer estimaciones no sesgadas de asociación en estudios observacionales que incluyen el control y el ajuste por efectos de confusión y modificación de efecto.

Aquí nos limitaremos a recapitular los aspectos esenciales de la evaluación de hipótesis de asociación dentro del marco de un estudio de cohorte analítica.

En primer lugar, la naturaleza y el nivel de medición de la variable principal de desenlace definen cuál tipo de medida de resumen es el apropiado para contrastar los resultados entre los diferentes grupos de exposición.

El desenlace de interés puede ser una variable continua que se mide en una escala numérica de razón o de intervalo; por ejemplo, niveles de tensión arterial media, glucemia, carga viral, índice de masa corporal (IMC), etc. En muchos casos se escoge como variable de desenlace un evento discreto, que suele tener más significación y mayor relevancia clínica: desarrollo de hipertensión arterial, episodios de hipoglucemia, remisión o curación microbiológica, control del sobrepeso, etc. Estas variables discretas de interés pueden

ser binarias: muerte/sobrevida, infección o secuelas, presente o ausente; o bien, pueden haber sido graduadas en una escala ordinal (leve, moderado, severo). El desenlace también puede expresarse como conteos de eventos, absolutos (número de episodios de hipoglucemia) o relativos (tasa de episodios de hipoglucemia por tiempo-paciente), y, finalmente, puede tratarse del tiempo hasta la ocurrencia de eventos.

La estrategia y las técnicas específicas de análisis deberán ajustarse a la naturaleza y la escala empleada para medir las variables de desenlace.

En la discusión que sigue sobre el enfoque general del análisis, se asume (como suele ser el caso en muchas investigaciones) que el desenlace de interés es ***discreto dicotómico***. El estudio de cohorte permite, entonces, estimar, en una muestra de una población de prueba, la asociación entre la exposición a diferentes grados de un factor de interés (riesgo, pronóstico, causal) y la frecuencia o la intensidad de al menos una variable de desenlace. Permite también comparar la frecuencia relativa de presentación del desenlace entre pacientes expuestos y no expuestos, o entre grupos con diferentes grados de exposición. La **figura 17.6** ilustra la manera de calcular el RR, que es el cociente entre las incidencias acumuladas (proporciones de incidencia de un evento discreto dicotómico) observadas en al menos dos niveles de exposición al factor de interés.

Otras maneras de resumir y contrastar las frecuencias de los desenlaces de interés incluyen la diferencia de riesgos (***riesgo atribuible***), la fracción de diferencia de riesgos (***fracción de riesgo atribuible***) y el número necesario para tratar o para exponer (véase el **capítulo 9** "Medidas de frecuencia, de asociación y de impacto").

Inicialmente se hace un cálculo del RR "crudo" comparando las incidencias halladas por el estudio. Dicho estimador podría verse afectado por distintas fuentes de error sistemático, tal como se ha venido discutiendo a lo largo del presente capítulo. Las precauciones tomadas en el diseño y la conducción del estudio permiten anticipar y controlar muchas de esas fuentes de error. Hay, sin embargo, una fuente que siempre amenaza la validez del estimador de asociación obtenido de un estudio observacional analítico, y es la acción de los modificadores de efecto y de los factores de confusión.

Durante el diseño se debieron identificar los potenciales modificadores de efecto y de factores de confusión. Algunos se controlaron, por ejemplo, mediante exclusión o emparejamiento. En la mayoría de los casos, no es eficiente o factible controlar más de dos o tres factores a través de dichas estrategias. Se debe, entonces, cuantificarlos y controlarlos en el análisis.

Inicialmente se evalúa cuáles factores candidatos tienen el verdadero potencial de confundir y sesgar las estimaciones. Cabe recordar que un factor puede confundir si se asocia a la variable de desenlace, si se distribuye diferencialmente con la variable de exposición (o sea, se asocia a ella), y si no forma parte de la vía causal postulada entre la exposición y el desenlace. Para evaluar el potencial de confusión, en primer lugar se examina la distribución de los factores candidatos entre los grupos de pacientes expuestos y no expuestos. Un factor que se distribuye balanceadamente entre grupos de exposición no confunde, y se puede dejar de tenerlo en cuenta en el análisis. Los factores con distribución diferencial deben contrastarse con el desenlace. Si muestran asociación al desenlace, pueden haber distorsionado (confundido) la asociación entre exposición y desenlace. Si ese es el caso, se deben usar técnicas estadísticas, tales como análisis estratificados y modelos multivariados, para cuantificar y remover el efecto de los factores de confusión y obtener un estimador ajustado

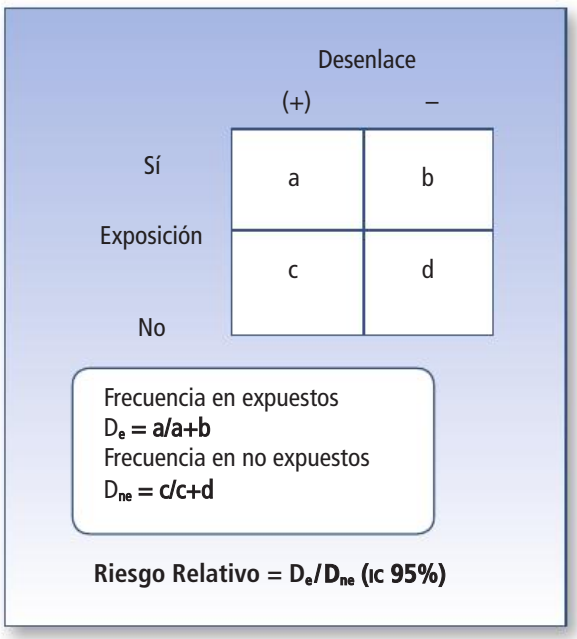


FIGURA 17.6 **Comparación de incidencias entre pacientes** expuestos y no expuestos, en un estudio analítico longitudinal, mediante el cálculo del riesgo relativo (razón de riesgo).

(no sesgado) de la asociación. En los **capítulos 16 "Estudios de casos y controles", 17 "Estudios** de cohorte analítica" y **29 "Plan de análisis"), se** discute en detalle el uso apropiado de dichas técnicas, incluyendo la evaluación y la cuantificación de la modificación de efecto (***interacción*).

CONCLUSIÓN

Los estudios de cohorte analítica son una herramienta de gran utilidad para evaluar hipótesis de asociación en los casos donde una asignación experimental en seres humanos es imposible, impráctica o contraria a la ética. La arquitectura del estudio, que evalúa la frecuencia de aparición de un desenlace en grupos con diferentes niveles de exposición, permite establecer si la exposición precede al desenlace, y si la aparición del desenlace es diferencial entre grupos con distintos niveles de exposición. Como todos los estudios analíticos observacionales, el diseño es vulnerable a sesgos que se asocien a la asignación observacional.

El diseño de cohorte resulta eficiente para estudiar condiciones relativamente comunes y con periodos de latencia relativamente cortos. Cuando se trata de condiciones poco comunes o de periodos de latencia muy prolongados, otras estrategias, tales como los estudios de casos y controles, pueden ser mejores opciones, y, en ocasiones, la única opción viable.

Un estudio de cohorte permite estimar de forma válida riesgos de aparición de eventos (proporciones de incidencias), tiempos de supervivencia, tasas y densidades de incidencias, lo que no se puede hacer con los estudios de casos y controles, excepto en condiciones especiales en algunos diseños anidados.

También permite evaluar, a partir de una exposición de interés, múltiples posibles desenlaces; incluso, los de enfermedades diferentes que solo tengan en común los factores de riesgo que se las asocian: por ejemplo, la enfermedad pulmonar obstructiva crónica (**EPOC**), el cáncer pulmonar y otras neoplasias, la enfermedad cardiovascular isquémica, la enfermedad cerebrovascular son todos desenlaces que pueden ser evaluados en un estudio de cohorte analítica sobre los potenciales efectos del tabaquismo sobre la salud.

Hay múltiples variantes del estudio de cohorte analítica, definidas de acuerdo con su marco temporal (históricas ***vs.* concurrentes),** asignación ***vs.*** selección (una cohorte ***vs.* múltiples** cohortes), el manejo de la exposición y del tiempo de exposición (análisis tradicional ***vs.* panel o *vs.* modelos longitudinales), e, incluso,** según el planteamiento de la evaluación de la asociación: longitudinal, ***vs.* anidar un** estudio de casos y controles en una cohorte. Más aún, hay diseños híbridos (caso-cohorte) o de controles no concurrentes (cohorte expuesta concurrente ***vs.* cohorte histórica no expuesta),** y hasta múltiples cohortes históricas en tiempos diferentes, pero sucesivos (análogos a las series temporales interrumpidas).

En la evaluación de causalidad, el estudio de cohorte ha sido, tradicionalmente, la piedra angular para hallar la evidencia de asociación causal entre factores de riesgo y muchas de las enfermedades crónicas no transmisibles, e, incluso, lesiones por accidentes, que explican una gran parte de la carga de enfermedad que actualmente experimentan la población pediátrica mayor y la población adulta en el mundo. Los estudios básicos, en modelos animales, los reportes de casos y los estudios de casos y controles suelen preparar el escenario, pero la prueba reina de causalidad se construye casi siempre con estudios prospectivos del tipo de la cohorte analítica.

En general, los estudios de cohorte involucran tamaños de muestra más elevados que los de casos y controles, así como tiempos de seguimiento más prolongados; también consumen muchos más recursos y se demoran más para ofrecer respuestas. Por tal motivo, si el investigador decide que la mejor manera de contestar su pregunta de investigación es acudir a esta herramienta, poderosa, pero costosa (la "artillería pesada" de la investigación de riesgo, pronóstico y causal en humanos), deberá asegurarse de que la planeación, el diseño, la conducción y el análisis son óptimos.

Los errores que invaliden los resultados de un estudio de cohorte son extremadamente costosos, en términos no solo monetarios, sino de conocimiento y de provisión de atención en salud; y la sociedad no se puede dar el lujo de permitir que sus investigadores los cometan.

Lecturas recomendadas

Greenberg R, Daniels S, Flanders D. Cohort studies. En: Greenberg R, Daniels S, Flanders D, editors. Medical epidemiology. New York: Lange; 2001.

Kelsey JL, Whittemore AS, Evans AS, et al. Cohort studies, statistical analysis I. En: Kelsey JL, Evans AS, Thompson WD, editors. Methods in observational epidemiology. 2nd ed. New York: Oxford University Press; 1996. p. 130-66.

Kelsey JL, Whittemore AS, Evans AS, et al. Cohort studies, statistical analysis II. En: Kelsey JL, Evans AS, Thompson WD, editors. Methods in observational epidemiology. 2nd ed. New York: Oxford University Press; 1996. p. 167-87.

Kelsey JL, Whittemore AS, Evans AS, et al. Panel studies. En: Kelsey JL, Evans AS, Thompson WD, editors. Methods in observational epidemiology. 2nd ed. New York: Oxford University Press; 1996. p. 2578.

Kelsey JL, Whittemore AS, Evans AS, et al. Prospective cohort studies: Planning and execution. En: Kelsey JL, Evans AS, Thompson WD, editors. Methods in observational epidemiology. 2nd ed. New York: Oxford University Press; 1996. p. 86-112.

Londoño JL. Estudios de supervivencia. En: Londoño JL, editor. Metodología de la investigación epidemiológica. Medellín: Editorial Universidad de Antioquia; 1995. p. 193-216.

Rothman KJ, Greenland S, Lash TL. Precision and validity in epidemiological studies. En: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. 2nd ed. Boston: Lippincott-Raven Publishers; 1998. p. 126-33.

Rothman KJ, Greenland S, Poole C, et al. Causation and causal inference. En: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. 2nd ed. Boston: Lippincott-Raven Publishers; 1998. p. 7-16, 23-8.

CAPÍTULO

18 Estudios sobre correlación y concordancia

Ajustar a alto de página

LUIS FELIPE URIZA CARRASCO
MARÍA NELCY RODRÍGUEZ MALAGÓN

Aunque todos los expertos estén de acuerdo, pueden estar equivocados.

BERTRAND RUSSELL

INTRODUCCIÓN

Predecir una variable si se conoce otra, o decidir si una prueba diagnóstica puede o no reemplazar a una que está en uso son preguntas habituales en medicina, y a las que se responde mediante estudios y pruebas de correlación y concordancia.

A menudo, en la evolución de la medicina se deben evaluar nuevas tecnologías o nuevas técnicas para medir variables fisiológicas, químicas o anatómicas, entre otras. La primera pregunta que nos hacemos es si la nueva técnica supera a la anterior o si es igual, pues casi siempre en el desarrollo de pruebas diagnósticas se busca superar las existentes, bien porque tengan mejor desempeño diagnóstico, o bien, porque ofrezcan mejores resultados en cuanto a costos, disponibilidad, automatización, rapidez, ergonomía o simplicidad, entre otros aspectos.

Podemos comparar, entonces, las cualidades del nuevo examen por las características operativas, comparar su sensibilidad y su especificidad, comparar curvas de características operativas (**ROC** ***curve***, por las siglas en inglés de ***receiver operating characteristics***) o hacer análisis económicos; sin embargo, no siempre se puede acceder al patrón de oro para calcular sensibilidad y especificidad, y, a veces, las circunstancias éticas pueden hacer imposible que se tome una biopsia de un órgano o se haga una intervención quirúrgica. En esos casos, para tomar la decisión sobre si el nuevo modelo de examen debe usarse, y si puede sustituir al antiguo en la práctica, debe demostrarse que dicha prueba es confiable, y solo en ese contexto es adecuado evaluar la concordancia.



Para evaluar el acuerdo entre dos mediciones, se dispone de varias opciones; entre ellas, la proporción de aciertos. Sin embargo, si en una enfermedad muy común se mide una variable usando dos métodos, ambos tendrán probabilidades de tener aciertos debidos solo al azar.

Por otra parte, la proporción de acuerdos es difícil de medir cuando las variables son continuas o están categorizadas.

EJEMPLO

Se quiere estudiar la relación entre el consumo alto de grasas y el de alcohol con la ocurrencia de pancreatitis; o el acuerdo entre las interpretaciones de un estudio radiológico hechas por dos radiólogos para el diagnóstico de una enfermedad del tracto respiratorio; o los valores esperados en los parámetros de la función pulmonar a partir de las características físicas de un individuo, como su edad, su peso o su estatura.

En cualquiera de las anteriores circunstancias se puede pensar que en los estudios planteados se evaluará tan solo la relación entre las variables. Sin embargo, cada una de las situaciones descritas involucra un análisis estadístico distinto, según el objeto de estudio. Para el caso del estudio de la pancreatitis, se debe buscar la asociación entre las variables, para lo que es necesario buscar una medida que hable de la dirección y de la magnitud de dicha asociación. En el caso de la interpretación de un estudio radiológico por dos radiólogos, nos vemos ante la necesidad de evaluar la concordancia o el acuerdo entre esas observaciones, y, por último, en el caso de determinar el valor esperado del parámetro de la función pulmonar para un sujeto, nos hallamos ante la necesidad de predecir una variable a partir de una o más características que se constituyen en factores independientes en el estudio[1].

En el presente capítulo se consideran algunos otros ejemplos que ilustran, especialmente, las diferencias entre situaciones donde debería considerarse un análisis de correlación, y otras en las que se debe considerar el análisis de concordancia.

Antes, sin embargo, es importante tomar en cuenta los conceptos de ***validez*** y de ***confiabilidad***, que ya han sido tratados a profundidad en otros capítulos.

MÉTODOS PARA LA EVALUACIÓN DE LA CONCORDANCIA

Cuando el interés reside en evaluar en qué grado dos o más métodos coinciden en cuanto a sus resultados en relación con la observación sobre un mismo individuo, o cuando uno o más observadores evalúan a un mismo sujeto, estamos frente a un problema de concordancia.

En tales casos puede ocurrir que uno de los métodos sea considerado un "patrón de oro" o un estándar. Si este es el caso, a la concordancia se la llama ***conformidad***. Por el contrario, si ninguno de los métodos puede ser considerado patrón de oro o referencia, la concordancia encontrada será una medida de ***consistencia***[1].

Según el tipo de escala de medición empleada para observar el evento de interés, las medidas o los índices de concordancia que reflejan el grado de acuerdo pueden variar.

Así, deberá usarse una medida estadística cuando el evento de interés es de tipo nominal (dicótomo o no); otra distinta, cuando tal evento se mide recurriendo a una escala ordinal, y otra, cuando el evento es de tipo numérico (y, especialmente, si es continuo).

A continuación, se revisarán algunas situaciones para ilustrar la determinación y el tipo de cálculo de las medidas de concordancia según el tipo de evento estudiado.

Evaluación de la concordancia en el caso de variables categóricas

Variables o eventos nominales

En el ejercicio de las ciencias de la salud, es indispensable alcanzar la máxima confiabilidad.

Y para ello, uno de los aspectos importantes que se deben evaluar es la ***concordancia***, o acuerdo en las observaciones realizadas cuando estas son tomadas por diferentes obervadores o métodos, sobre un mismo sujeto. Este acuerdo puede ser medido entre el mismo observador (***concordancia intraobservador***) o entre diferentes observadores (***concordancia interobservador***).

Aunque virtualmente todas las observaciones clínicas tienen algún grado de inconsistencia, la mayoría son, de forma notoria, más consistentes que lo esperado tan solo por efecto del azar.

Hay muchas formas de establecer dicha concordancia. La más elemental es calcular la proporción de casos donde el observador concuerda consigo mismo o con otro observador con entrenamiento similar. Antes de comparar, es necesario tener en cuenta que los errores asociados a la calificación de los observadores son independientes, y ello requiere que los pacientes o los sujetos sean independientes. La calificación de un observador no debe influir en la de otro.

Para comprender lo anterior, analicemos el siguiente ejemplo hipotético: Dos radiólogos valoraron los resultados de 100 radiografías de tórax, y las clasificaron como "normales" o "anormales". Las valoraciones se hicieron de forma independiente, así que ninguno sabía la opinión del otro sobre la muestra analizada **(tabla 18.1)**.

El radiólogo 1 hizo el diagnóstico de anormal en 30 casos, y el segundo lo hizo en 35 casos. En 20 casos hubo acuerdo total en cuanto al diagnóstico de anormal. Un abordaje simplista diría que el grado de acuerdo fue del 20 %. Sin embargo, dicha cifra estaría desconociendo que hubo 55 casos donde ambos observadores coincidieron en considerar "normales" los estudios. Decir que la concordancia total fue del 75 % (20 % + 55 %) es sobreestimar el acuerdo, por desconocer que, como solo había 2 categorías diagnósticas posibles ("normal" y "anormal"), algún grado de coincidencia se daría solo por azar. Aquí, entonces, debemos recurrir a una prueba estadística que permita medir el grado de concordancia entre dos observadores ante una variable dicótoma y que no sea atribuida al azar. Dicha prueba es llamada Kappa (**κ, por la** letra griega de ese nombre)[2]. La fórmula para su cálculo (relativamente sencillo) es:

$$k = \frac{o - c}{1 - c}$$

Donde ***o* es la concordancia observada, y *c*,** la concordancia por azar.

Para el ejemplo que venimos desarrollando, los cálculos serían así:

- Se calcula en cuántas de las radiografías donde hay concordancia, esta se debe al azar, lo cual, a su vez, se determina mediante la multiplicación de cada una de las cantidades encontradas positivas (para el ejemplo, serían las cantidades en las que se diagnosticó como "anormal"), y se divide el resultado entre el número total de radiografías analizadas, que en este caso son 100. Tenemos, pues: (30 × 35) / 100 = 10,5 %.
- El segundo paso es calcular en cuántas de las radiografías donde hubo concordancia con el diagnóstico negativo (es decir, "normal"), esta fue por azar, lo cual, a su vez, se determina multiplicando cada una de las cantidades halladas como negativas, y dividiendo el resultado entre el total de radiografías analizadas. Tenemos así: (70 × 65) / 100 = 45,5 %.
- El tercer paso consiste en sumar las cantidades halladas anteriormente, y dividirlas entre 100. Con eso obtenemos una proporción para la concordancia debida al azar. El resultado es: (10,5 + 45,5) / 100 = 0,56, o 56%.
- La concordancia observada (***o*) es, entonces:** 20 % + 55 % = 75 %, o 0,75; por lo tanto, la concordancia más allá del azar es: 0,75 – 0,56 = 0,19; esta cifra será el numerador de **κ.**
- El potencial de concordancia más allá del azar (1 – c) es, simplemente, el 100 % menos la concordancia por azar; el resultado es: 100 % – 56 % = 44 %, o 0,44. Dicha cifra constituye el denominador de **κ; por lo tanto,** tenemos, finalmente, esta ecuación:

$$\kappa = 0{,}19 / 0{,}44 = 0{,}43.$$

Tabla 18.1 Datos de concordancia de la interpretación de radiografías por dos radiólogos

		Radiólogo 1		
		Anormal	**Normal**	**Total**
Radiólogo 2	Anormal	20	15	35
	Normal	10	55	65
	Total	30	70	100

La forma de interpretar estos resultados corresponde a la proporción de acuerdo mayor que el esperado por el azar, que para estos 2 radiólogos fue de 0,43; o sea, la concordancia más allá de la esperada por efecto de azar fue del 43 %.

Para clasificar los resultados de un índice de κ, se han establecido puntos de corte de manera arbitraria, y la más utilizada es la escala descrita por Landis y Koch[2], así:

- *De 0-0,2*: Débil.
- *De 0,2-0,4*: Leve
- *De 0,4-0,6*: Moderada.
- *De 0,6-0,8*: Sustancial, o alta.
- *De 0,8-1,0*: Casi perfecta.

En esta escala, los valores de κ siempre son positivos, y el rango es de 0-1. Tiene el inconveniente de que se superponen valores en el límite. El valor puntual, sin embargo, no expresa todo, y debe ir acompañado de un intervalo de confianza (IC), que se calcula con la siguiente fórmula para el 95 %, con la constante 1,96:

$$k - 1{,}96 \times SEk \text{ hasta } k + 1{,}96 \times SEk$$

El error estándar para el κ se calcula con la siguiente fórmula:

$$SEk = \frac{p(1-p)}{\sqrt{n(1-p2)^2}}$$

En nuestro caso, el valor del error estándar del κ, teniendo en cuenta que $P = 0{,}75$ y $Pe = 0{,}56$ corresponde a 0,098. Entonces, el IC es:

$$(0{,}43 - 1{,}96 \times 0{,}098 \text{ hasta } 0{,}43 + 1{,}96 \times 0{,}098),$$
$$\text{o } (0{,}24\text{-}0{,}62).$$

Vemos, entonces, que nuestro valor se encuentra en un amplio IC, que incluye un rango de interpretación desde *leve* hasta *sustancial*. Dicho IC nos indica que es deseable, en nuestro ejemplo, aumentar el tamaño de muestra; sin embargo, el 95 % de los IC incluyen el valor encontrado.

Otro valor que nos puede ayudar en la interpretación es el valor de *p*, que en nuestro ejemplo es < 0,05. La prueba de hipótesis, que, por defecto se calcula a una cola, es que el κ será > 0 (y en nuestro ejemplo lo es). Sin embargo, se pueden diseñar pruebas de hipótesis para encontrar valores de κ superiores a un valor previamente estipulado. En este caso, si nuestra hipótesis fuera encontrar un valor de concordancia casi perfecta o superior a 0,8, no habría podido corroborarse con nuestro estudio.

El valor de κ está influido por la prevalencia de la condición clínica o el atributo que se está evaluando[3,4].

La verdadera prevalencia de una condición es la que afecta el cálculo de dicha estadística, pero puede lograrse una medida indirecta usando el *índice de prevalencia*, que se puede calcular así:

$$\text{Índice de prevalencia} = \frac{|a-d|}{n}$$

En nuestro ejemplo, el índice de prevalencia tiene un valor de 0,35.

Si se utilizan 100 radiografías de otro hospital y se las clasifica también como normales o anormales (por parte de 2 radiólogos), tenemos los resultados que se presentan en la **tabla 18.2.**

El índice de prevalencia para esta situación es de 70 – 5 / 100 = 0,69.

Según estos datos, la concordancia esperada por el azar aumentó a 0,71, y, entonces, el cálculo del coeficiente κ sería:

$$\kappa = 0{,}75 - 0{,}71 \,/\, 1 - 0{,}71 = 0{,}03 \,/\, 0{,}29 = 0{,}14.$$

En el ejemplo, vemos cómo al aumentar la prevalencia, la probabilidad de acuerdo por el azar es mayor, y entonces el acuerdo más allá del azar es más bajo.

Es de importancia, pues, saber el efecto de la prevalencia en nuestro cálculo y en la interpretación.

Otro factor que puede afectar el cálculo, y, por ende, la interpretación del coeficiente κ, es el ejercicio de quién clasifica los resultados en cada categoría. Si un observador está adivinando o si clasifica los desacuerdos con preferencia hacia una de las categorías, también puede afectar más allá del azar el resultado del acuerdo.

Este efecto también se puede medir con un *índice de sesgo*, de la siguiente forma:

$$\text{Índice de sesgo} = \frac{|b-c|}{n}$$

En nuestro ejemplo, el valor del índice es de 0,05.

Tabla 18.2 Datos de concordancia de la interpretación de radiografías por dos radiólogos

		Radiólogo 1		
		Anormal	Normal	Total
Radiólogo 2	Anormal	70 (a)	15 (b)	85
	Normal	10 (c)	5 (d)	15
	Total	80	20	100

En otro ejemplo, con dos observadores menos experimentados, se obtuvo la siguiente información **(tabla 18.3)**:

En esta situación, el índice de sesgo es 0,15.

El acuerdo esperado por el azar teniendo en cuenta nuestro ejemplo inicial cambió a 0,51, y el cálculo del coeficiente κ sería el siguiente:

$$\kappa = 0{,}75 - 0{,}51 / 1 - 0{,}51 = 0{,}24 / 0{,}49 = 0{,}49.$$

Si bien en este caso la interpretación de acuerdo con la escala de Landis y Koch no cambia, sigue siendo una concordancia moderada. En otras situaciones, y cuanto mayor sea el sesgo, mayor será el aumento del coeficiente κ, y es importante anotar que esto puede llevar a cambios en la interpretación, de tal forma que para interpretar el resultado es importante tomar en cuenta la distribución de los desacuerdos.

κ *ajustado*

Aunque se ha propuesto corregir el cálculo del κ tanto por prevalencia como por sesgo, aún hay mucha discusión al respecto, y algunos autores piensan que el efecto de los anteriores es favorable para el κ, y que el investigador debe interpretar dicha estadística teniéndolos en cuenta[5]; otros creen que el sesgo y la prevalencia penalizan el κ de manera adecuada, pero debe informarse no solo el valor de κ, sino presentar los datos suficientes, para que sea posible buscar esos efectos[4].

Variables o eventos de tipo ordinal

Para situaciones en las que el interés sea evaluar la concordancia entre observadores o evaluar métodos mediante observaciones provenientes de eventos medidos en escalas ordinales, se debe considerar el uso del coeficiente κ *ponderado*. El cálculo de este índice es más complicado que el *kappa* simple, y se aplica para evaluar la concordancia cuando los resultados *no son dicótomos*. En esta modalidad, se permite medir cuán lejos de la "verdad" estaba cada observador.

Tabla 18.3 Datos de concordancia de la interpretación de radiografías por dos radiólogos

		Radiólogo 1		
		Anormal	Normal	Total
Radiólogo 2	Anormal	20 (a)	5 (b)	25
	Normal	20 (c)	55 (d)	75
	Total	40	60	100

Tabla 18.5 Resultado del κ no ponderado para el ejemplo de la tabla 18.4

κ observado	Error estándar	IC 95 %	IC 95 %
0,1283		Límite inferior	Límite superior
Método 1	0,0406	0,0487	0,2079
Método 2	0,0384	0,0531	0,2035

Tabla 18.6 Ponderación lineal de las diferencias

		Radiólogo			
Residente RI		T1	T2	T3	T4
	T1	0	1	2	3
	T2	1	0	1	2
	T3	2	1	0	1
	T4	3	2	1	0

Tabla 18.7 Ponderación cuadrática de las diferencias para el mismo ejemplo de la tabla 18-4

		Radiólogo			
Residente RI		T1	T2	T3	T4
	T1	0	1	4	9
	T2	1	0	1	4
	T3	4	1	0	1
	T4	9	4	1	0

Tabla 18.8 Para el mismo ejemplo, este es el resultado del κ con ponderación lineal y con su

κ observado	Error estándar	IC 95 %	IC 95 %
0,2284		Límite inferior	Límite superior
Método 1	0,0368	0,1563	0,3005

Tabla 18.9 Resultado del κ con ponderación cuadrática y con su , para el mismo ejemplo

κ observado	Error estándar	IC 95 %	IC 95 %
0,3518		Límite inferior	Límite superior
Método 1	0,0404	0,2726	0,431

de los desacuerdos y utilizando la ponderación cuadrática, que es la más recomendada; el acuerdo más allá del azar es moderado.

Como se puede ver en el ejemplo, el coeficiente mejoró, en apariencia, con el grado de formación del residente.

Pero igual a lo mencionado para la interpretación del **κ simple, no se deben olvidar los posibles sesgos,** y en este caso se debe considerar la diferencia en las frecuencias de la enfermedad en las categorías antes de atribuir el mayor acuerdo a un mejor grado de conocimiento del residente de segundo año.

Tabla 18.10 Datos de concordancia de la interpretación de tomografías de tórax por un radiólogo y un residente de segundo año, para la adecuada categorización del cáncer de pulmón

		Radiólogo 1				
		T1	T2	T3	T4	Total
Residente RI	T1	14	4	7	1	26
	T2	2	40	12	4	58
	T3	7	10	14	9	40
	T4	4	3	8	21	36
	Total	27	57	41	35	160

Tabla 18.11 Resultado del κ no ponderado para el ejemplo de la tabla 18-10, y con su IC

κ observado	Error estándar	IC 95 %	IC 95 %
0,3923		Límite inferior	Límite superior
Método 1	0,0538	0,2869	0,4977
Método 2	0,053	0,2884	0,4962

Tabla 18.12 Resultado del κ con ponderación cuadrática y con su IC, para el mismo ejemplo

κ observado	Error estándar	IC 95 %	IC 95 %
0,469		Límite inferior	Límite superior
Método 1	0,0849	0,3026	0,6354

Además de la prevalencia, hay otras condiciones que influyen en el cálculo y en la interpretación, como el número de categorías, pues a mayor número de categorías será más difícil lograr una clasificación correcta, lo que puede causar valores de κ más bajos. Otro factor es la subjetividad que se introduce con la ponderación; por último, cuando se construyen categorías a partir de variables continuas, el valor del κ depende de las definiciones arbitrarias de esas categorías.

Dichas limitaciones hacen que no sea recomendable comparar el valor de κ en poblaciones diferentes o si las categorías no son las mismas[6,7].

En resumen, si se quiere evaluar el acuerdo entre diagnósticos sobre la presencia o no de un tumor, la presencia o no de una neumonía o una bronconeumonía (***eventos de tipo nominal***), la medida indicada para determinar ese acuerdo es el coeficiente κ. En caso de que el acuerdo buscado sea sobre eventos de tipo ***ordinal*** (como, por ejemplo, la clasificación por grados de pancreatitis, trauma en órganos sólidos o la clasificación de los estadios sobre algún tipo de cáncer), la medida más indicada es el coeficiente **κ ponderado.**

Medidas de concordancia para el caso de variables numéricas

Cuando se tiene interés en determinar y evaluar el nivel de acuerdo entre diferentes métodos o entre dos o más observadores en relación con una característica de tipo numérico, se dispone de algunas medidas distintas del coeficiente κ, o κ ponderado. Cabe resaltar que la principal utilidad de esta medición es determinar si los métodos de evaluación son intercambiables, o si los diferentes observadores son igualmente confiables.

EJEMPLO

Puede haber interés en medir si puede utilizarse la glucometría (medición de glucosa por punción digital) en reemplazo de la glucemia medida en sangre venosa, o si la medición de la longitud de una extremidad por parte de cualquiera de los ortopedistas es igualmente aceptable. Dichas medidas utilizan para su cálculo las desviaciones respecto a una medida de tendencia central como el promedio. Dos de estas medidas son el ***coeficiente de correlación intraclase*** (CCI) y una modificación de este, conocida como el *coeficiente de correlación intraclase de Lin*[8,9].

El CCI parte de un concepto básico, que fue introducido originalmente por Fisher, como una formulación especial del coeficiente ***r*** de Pearson, basado en un modelo de análisis de varianza. Sin embargo, se han encontrado dificultades para hacer una interpretación clínica correcta de los valores de este coeficiente, y, por algunas desventajas metodológicas, se han propuesto algunos métodos alternativos cuando se estudia la concordancia entre variables de tipo cuantitativo. Por ejemplo, Bland y Altman (1986)[10] propusieron un método gráfico que se basa en las diferencias individuales, que, además, permite hallar los límites de concordancia y hacer una exploración gráfica de las diferencias observadas que se hallen entre las características que se estudien.

En relación con el CCI, se tienen diferentes versiones, que pueden dar muy diferentes resultados cuando se aplican a un mismo conjunto de datos. Quienes investigan deben ser conscientes de estas diferencias en la forma de cálculo, y seleccionar la forma más apropiada para la información que se pretende analizar, según el diseño de investigación seleccionado y los objetivos del estudio. Como, en general, la forma básica de cálculo incluye un análisis de varianza, es importante determinar si para dicho análisis solo se tomará la forma más simple que se conoce, llamada ***varianza a una vía***, o si la forma como se tomaron las mediciones corresponden a la lectura de un observador, o si los datos corresponden al promedio de varios observadores[11].

Por otro lado, se debe reconocer que la interpretación fundamental del CCI es que este corresponde a una medida de la proporción de una varianza atribuible a los objetos de mediciones[12].

O también, a la proporción de la variabilidad total debida a la variabilidad entre las personas[13]. Los objetos mencionados pueden ser, además, radiografías, transparencias, fotografías, paquetes, etc. Las mediciones pueden ser simples lecturas (como, por ejemplo, pesos de recién nacidos, etc.)[12]. Este coeficiente se considera la medida más usada cuando se hace una evaluación de la repetibilidad de un grupo de mediciones; de ahí que se conozca como un ***índice de repetibilidad***[12]. Se diferencia del coeficiente de correlación clásico en que, por ejemplo, cuando se tienen dos métodos que se están evaluando, en su cálculo se incluye la diferencia promedio entre los dos métodos.

Una de las formas más usadas para el cálculo del CCI fue propuesto por Deyo *et al.*[13], como una modificación a la fórmula original propuesta por Kramer y Feinstein[1]. Esta expresión aplica cuando se tienen dos observaciones por cada sujeto; o sea, cuando se quiere hallar la medida de repetibilidad de dos métodos; por ejemplo, el método A y el método B:

$$ICC = \frac{A^2 + B^2 - C^2}{A^2 + B^2 + D^2 - C^2/n}$$

Donde:

A corresponde a la desviación estándar del método A.

B corresponde a la desviación estándar del método B.

C corresponde a la desviación estándar de las diferencias entre los dos métodos A y B.

D es el promedio de las diferencias entre los dos métodos.

N es el número de sujetos observados.

El resultado obtenido es un valor que varía entre -1 y 1.

A pesar de la forma sencilla del cálculo anterior, este CCI tiene algunas importantes desventajas, como lo mencionan Cepeda y Pérez[14]:

- Para calcular el CCI, se presume que los métodos provienen de una muestra de métodos, pero usualmente se tienen dos métodos fijos, que son los de interés en la comparación.
- Este coeficiente supone que el error de medición es similar en los métodos, lo cual, en muchos casos, no necesariamente es cierto.
- Es un coeficiente que depende de la variabilidad de los sujetos estudiados, lo cual no es conveniente, pues en la búsqueda de la

concordancia se quiere que los resultados sean independientes de los sujetos a los cuales se está evaluando.

Dadas las desventajas anteriores, Lin (1989) propuso el ***coeficiente de correlación de concordancia*** (CCC) o ***coeficiente de correlación intraclase*** (CCI) *de Lin*[9,15], el cual fue modificado por el mismo autor en el año 2000. Este coeficiente tiene la ventaja de considerar dos aspectos: la precisión y la exactitud. Da cuenta de la concordancia entre los datos observados y la línea de la perfecta concordancia, la cual, si fuese graficada, correspondería a la línea de 45° en un gráfico de dispersión entre los métodos.

Este coeficiente permite determinar qué tan dispersos se hallan los datos de la línea de perfecta concordancia; o sea, es un coeficiente que da cuenta tanto de la precisión en la estimación de la concordancia como del sesgo que ocurre en las observaciones.

La forma de cálculo es la siguiente[13]:

$$CCC = \frac{A^2 + B^2 - C^2}{A^2 + B^2 + D^2}$$

Donde, al igual que en el caso del CCI:

A corresponde a la desviación estándar del método A.

B corresponde a la desviación estándar del método B.

C corresponde a la desviación estándar de las diferencias entre los dos métodos A y B.

D es el promedio de las diferencias entre los dos métodos.

La estimación del CCI de Lin también se nota como ρ_c.

Debe tenerse en cuenta que el coeficiente CCI y el CCC se diferencian solo en el último término del denominador, término que en algunas ocasiones puede ser pequeño, aunque no siempre.

Además de lo anterior, es importante reconocer que el CCI de Lin puede ser descompuesto en dos partes: 1) una correspondiente a lo que se llama ***coeficiente de correlación r***, o *coeficiente de correlación de Pearson*, y el cual da cuenta de la precisión de la estimación del acuerdo basada en los datos, y 2) otra correspondiente al *factor de corrección del sesgo*; es decir, cuánto se distancian o se desvían las observaciones hechas por los observadores o por los métodos en relación con la línea de perfecto acuerdo.

Este coeficiente puede variar entre 0 y 1. Cuanto más cercano esté a 1,0, mayor será el nivel de concordancia entre las mediciones de los observadores o entre los resultados arrojados por los dos métodos que se están comparando.

McBride (2005)[16], propuso puntos de corte para el CCI de Lin cuando se está evaluando el grado de acuerdo entre dos métodos. Esta es una escala desarrollada con criterios similares a los de la obtenida por Landis y Koch[2]:

- ***Acuerdo casi perfecto***: > 0,99.
- ***Acuerdo sustancial***: 0,95-0,99.
- ***Acuerdo moderado***: 0,90-0,95.
- ***Acuerdo pobre***: < 0,90.

No obstante lo anterior, cabe anotar que si el resultado del coeficiente fuese de 0,95 podría considerarse moderado o sustancial. Para efectos de una correcta interpretación, debería dejarse el nivel de acuerdo moderado entre 0,90-0,949.

EJEMPLO

Los siguientes son los resultados de un estudio hipotético, en el cual se tomaron mediciones de la presión intraocular de 40 pacientes, mediante 2 diferentes métodos A y B. La **figura 18.1** muestra los resultados de las mediciones de los 2 métodos en relación con la línea de perfecto acuerdo. La estimación puntual y por intervalo del CCI de Lin fue ρ_c = 0,941; IC 95 %: (0,91-0,97). Este valor, según McBride[16], corresponde a un nivel de concordancia moderado.

Para los mismos datos se usa la ***gráfica de Bland y Altman***, que presenta la diferencia entre cada pareja de valores frente a la media de cada par. Cuando no se tienen errores sistemáticos en las mediciones, se espera que los puntos se distribuyan de forma aleatoria alrededor de la línea recta correspondiente a la diferencia de 0. La línea roja junto a la línea de 0 corresponde al promedio de las diferencias de cada par de valores observados. Las dos líneas punteadas (superior e inferior) son el límite superior e inferior de concordancia, respectivamente **(figura18.2).**

Diferencia entre la correlación y la concordancia

Relativamente a menudo, cuando se trata de comparar los resultados provenientes de dos o más métodos, o bien, datos resultantes

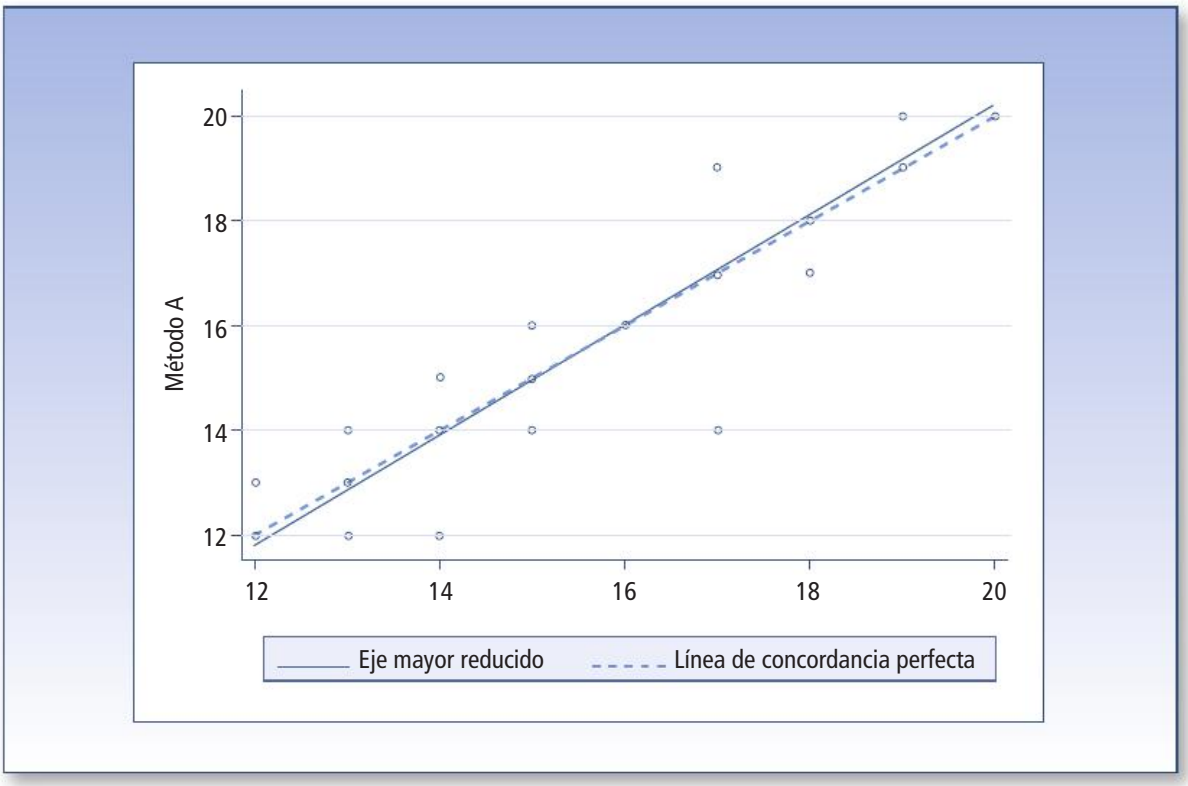


Figura 18.1 **Mediciones de los dos métodos en relación con la línea de perfecto acuerdo. Datos hipotéticos.**
Fuente: Elaboración de los autores.

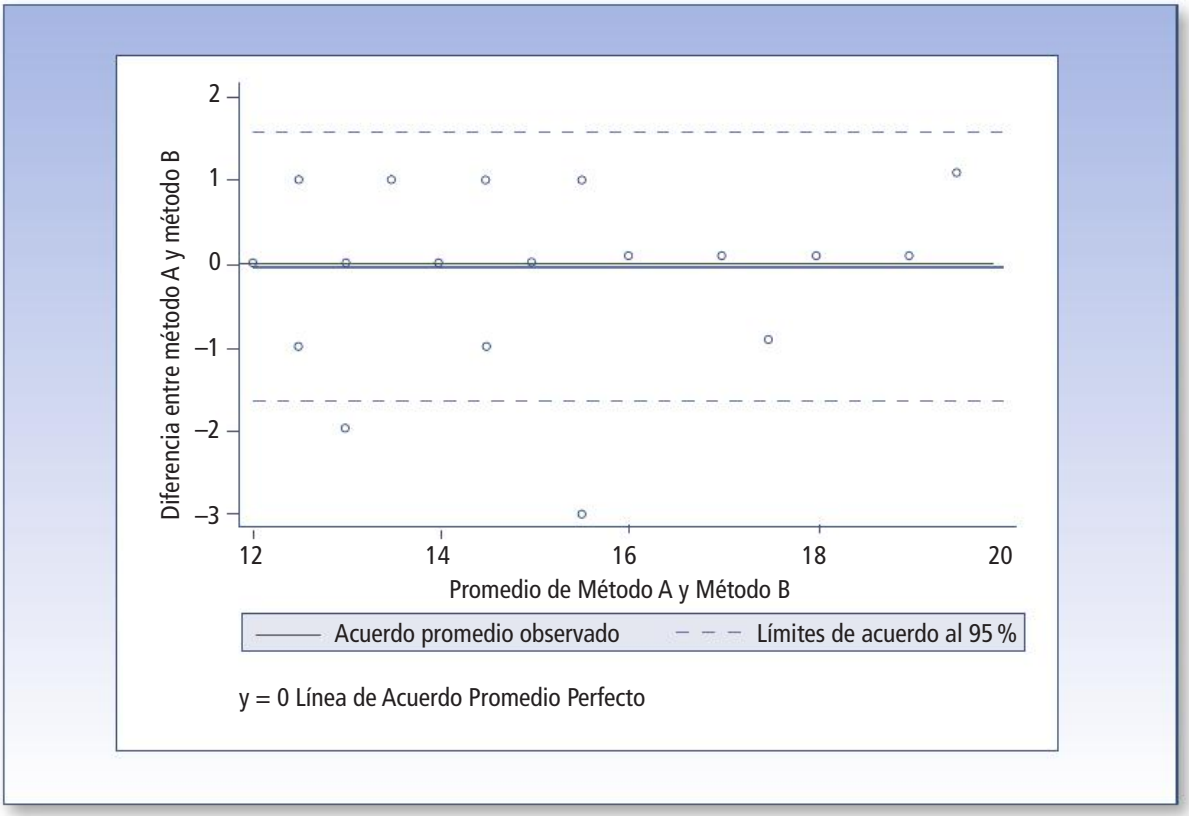


Figura 18.2 **Gráfica de Bland y Altman para los datos provenientes del método A y del método B.**

de la observación de dos o más observadores, y si las observaciones resultan en datos de tipo continuo, se utiliza el coeficiente de correlación de Pearson. Sin embargo, antes de usar dicho coeficiente, se debe revisar si el objetivo es ver cuál es la magnitud de la asociación entre los datos provenientes de los dos métodos u observadores (en cuyo caso estaríamos ante un problema de tendencia), o si el objetivo es ver qué tan concordantes son esos resultados de las observaciones obtenidas de los métodos o de los observadores sobre el mismo sujeto.

Es incorrecto usar correlación cuando se quiere saber si hay o no intercambiabilidad entre dos métodos.

EJEMPLO

Si se busca saber si un método para determinar la glucometría es o no confiable, el glucómetro puede tener una descalibración de 40 mg/dL, por lo que cada valor por glucometría será 40 mg más alto que la glucemia medida en la sangre venosa. Si se evalúa la correlación entre los dos resultados en una muestra de 150 pacientes, como quien tenga el valor más bajo en un método también lo tendrá en el otro método, y los valores más altos en una medición también corresponderán a los valores más altos de la otra medición, la correlación será perfecta o casi perfecta, aunque no haya coincidencia de ningún valor.

En ese caso, se concluiría, erróneamente, que los métodos son intercambiables, cuando, de hecho, la verdad es exactamente lo opuesto. En este caso, si se quiere saber si se pueden intercambiar los métodos, debe utilizarse la aproximación de medir la concordancia; idealmente, utilizando el ***índice de correlación y concordancia***.

Si se busca ver si hay una tendencia similar (***correlación lineal***) en la elevación de peso corporal y el aumento de los triglicéridos (vale decir, si cabe esperar que a mayor peso haya valores más altos de triglicéridos), la citada correlación es el método apropiado, y en este caso el índice de Pearson debería ser el método que se utilice.

Se puede hallar una correlación muy alta entre las observaciones, pero una concordancia pobre. Además, el coeficiente de correlación, denotado por ***r***, expresa el grado de tendencia *lineal* entre dos series de datos.

La forma de cálculo del mencionado coeficiente está dada por la siguiente expresión[13]:

$$r = \frac{A^2 + B^2 - C^2}{2AB}$$

Donde A, B y C tienen la misma definición dada para el cálculo del CCI.

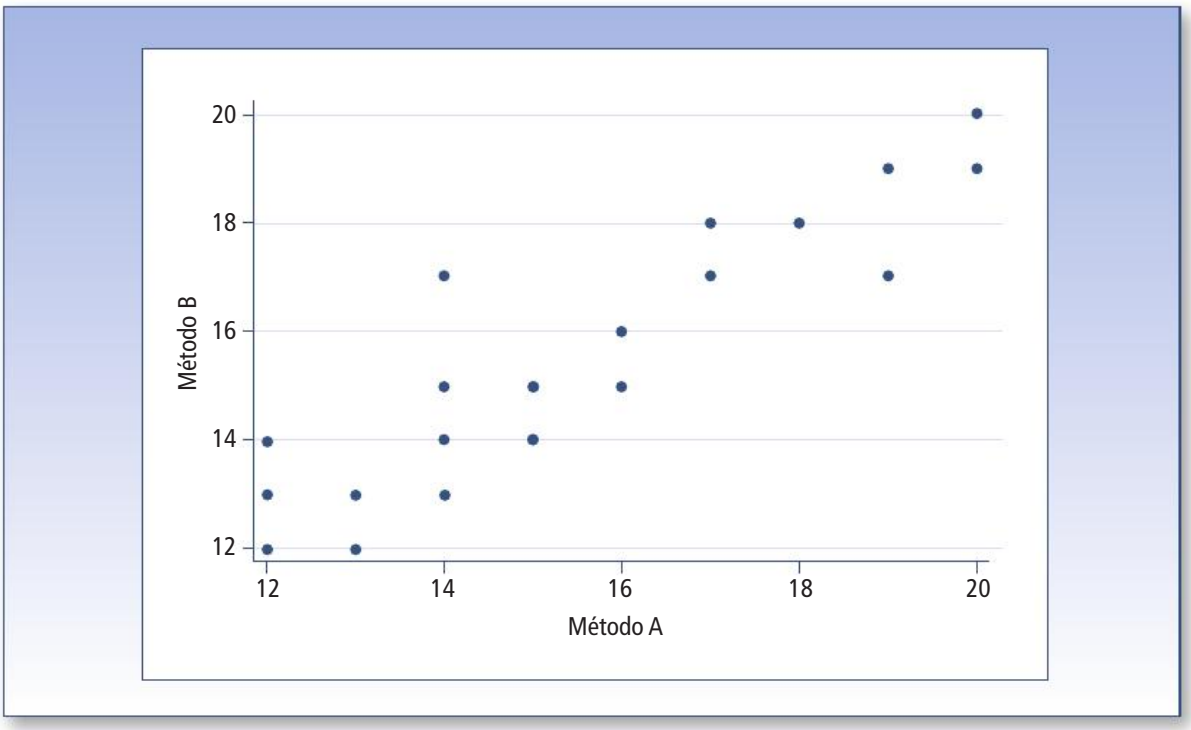


FIGURA 18.3 Diagrama de dispersión del método A *vs.* el método B.

Para el ejemplo de los datos sobre las mediciones de la presión intraocular por 2 métodos distintos, si calculamos el coeficiente de correlación, obtenemos $r = 0{,}9428$, lo cual nos dice que la magnitud de la asociación lineal entre el método A y el método B es del 94,28 %; ello indica una buena tendencia lineal, como se muestra en la **figura 18.3.**

Si bien en este ejemplo los valores hallados para el CCI de Lin y el valor del coeficiente de correlación de Pearson son casi iguales, su uso y su interpretación corresponden a la respuesta a dos preguntas diferentes. La una, relacionada con el acuerdo sobre dos métodos u observadores, y la otra, relacionada con la tendencia lineal de los dos métodos u observadores.

Referencias

1. Kramer MS. Alvan R. Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharm Ther. 1981;29:111-23.
2. Landis JR, Koch GC. The measurement of observer agreement for categorial data. Biometrics; 1977;33:159-74.
3. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43:543-9.
4. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551-8.
5. Thompson WD, Walters SD. A reapraisal of the Kappa coefficient. J Clin Epidemiol 1998;41:949-58.
6. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213-220.
7. Bland JM. University of York Department of Health Sciences Measurement in Health and Disease Cohen's Kappa [internet]. 2008 [citado 2008 jul]. Disponible en: https://www-users.york.ac.uk/~mb55/msc/clinimet/week4/kappash2.pdf.
8. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley & Sons; 1981.
9. Lin L. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45:255-68.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;I:307-10.
11. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability Psychol Bul. 1979;86:420-8.
12. Shoukri MM. Measures of Interobserver Agreement. Boca Ratón, FL: Chapman & Hall/CRC Press; 2010.
13. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Controlled Clin Trials. 1991;12:142S-158S.
14. Cepeda S, Pérez A. Estudios de concordancia: Intercambiabilidad en sistemas de medición. En: Ruiz A. Morillo E, Epidemiología Clínica. Investigación Clínica Aplicada. Bogotá: Ed. Médica Panamericana; 2004. Cap. 17.
15. Lin L. I-L. A note on the concordance correlation coefficient. Biometrics. 2000;56:324-5.McBride GB. A proposal for strength of agreement criteria for Lin´s concordance correlation coefficient. New Zealand: NIWA Project: MOh05201; 2005.
16. De Vet H. Observer reliability and agreement. En: Armitage P, Colton T, editores. **Encyclopedia** of Biostatistics. Vol 4. Chichester: John Wiley & Sons; 1998. p. 3123-7.

CAPÍTULO

19

Estudios cualitativos

María Beatriz Ospina García • Carlos Gómez-Restrepo

No todo lo que puede ser contado cuenta,
ni todo lo que cuenta puede ser contado.
Albert Einstein

INTRODUCCIÓN

Un paciente grave, con un infarto, en una Unidad de Cuidado Intensivo (UCI), con edema pulmonar y con múltiples complicaciones hemodinámicas y renales, recibe atención altamente especializada y vigilancia permanente. Deben vigilarse de forma estricta parámetros como el gasto cardíaco, la resistencia periférica, la densidad urinaria y los niveles séricos de ácido láctico. Y también deben vigilarse y controlarse su perfusión tisular, el dolor precordial, la ansiedad, la disnea, la depresión resultante de su condición, el insomnio y muchas otras variables que no parecen fáciles de medir, y que necesitan herramientas específicas para su evaluación, y, sobre todo, actitud y habilidades.

Si bien la división tiene más una utilidad metodológica, es claro que muchas de las variables más importantes en la atención en salud no pueden evaluarse mediante aparatos o exámenes de laboratorio. Y así como el pediatra debe poder interpretar el calcio y glucemia del recién nacido, también debe estar preparado para evaluar y controlar la ansiedad de la madre, o para evaluar el rendimiento escolar o el desarrollo psicosocial de su paciente. Y el cirujano, a su vez, necesitará herramientas para decidir si una nueva técnica quirúrgica mejora la percepción de dolor e incomodidad de su paciente, y si mejorará su calidad de vida.

Es claro, en la práctica diaria de las profesiones de la salud, que algunos de los aspectos más importantes de la atención pueden ser cuantificados de manera precisa, en tanto que otros, tanto o más importantes, son difíciles de definir, de cuantificar y de generalizar. Más que lo cuantitativo, a menudo es fundamental lo cualitativo.

¿Qué es la investigación cualitativa, y por qué es importante?

La investigación en salud utiliza a menudo abordajes cuantitativos, y al conocimiento obtenido mediante este enfoque se le considera objetivo, medible, reproducible y generalizable, pero también no pocas veces debe enfrentar los retos de fenómenos relacionados con los procesos de salud y enfermedad para los cuales no hay instrumentos de medición, o que son subjetivos y socialmente construidos; su significado es, necesariamente, personal y muy contextual, y requieren abordajes metodológicos diferentes. La investigación cualitativa es un componente metodológico esencial en la investigación en salud. Promueve la evaluación de problemas de salud con perspectivas diversas (tanto la

personal como la social y la cultural), y se vale de técnicas de recolección e interpretación de datos que favorecen el conocimiento de factores que no son directamente apreciables mediante el uso de métodos cuantitativos y ayudan a comprender el contexto donde suceden los procesos de salud, enfermedad, así como la organización de los servicios de salud.

La investigación cualitativa tiene una larga tradición en otras disciplinas como la educación, la sociología, la psicología y la antropología. Cuando se la aplica a la investigación en salud, es una poderosa herramienta metodológica para obtener respuestas profundas sobre lo que las personas piensan o sienten, y permiten una mejor comprensión de sus actitudes, sus creencias, sus motivos y sus comportamientos[1].

De manera general, los estudios cualitativos cumplen cuatro propósitos:

1. Como método principal de recopilación de datos en relación con un tema de investigación en temas de relevancia personal o social.
2. Generar ideas e hipótesis que con posterioridad puedan verificarse en estudios puramente cuantitativos.
3. Complementar los datos de un estudio predominantemente cuantitativo.
4. Apoyar la evaluación de un estudio cuantitativo.

La gran fortaleza de la investigación cualitativa es su habilidad para proporcionar información sobre aspectos intangibles, tales como normas sociales, papeles de género, aspectos relacionados con el origen étnico y la religión, entre otros, y cuyo papel en la atención en salud y en la investigación no siempre resulta evidente.

DIFERENCIAS ENTRE INVESTIGACIÓN CUALITATIVA E INVESTIGACIÓN CUANTITATIVA

Los abordajes cualitativos y cuantitativos en investigación difieren en su objetivo, en las preguntas que evalúan, en los instrumentos y las técnicas de recolección de datos que utilizan, en los datos que generan, en la manera como estos datos se analizan y en el grado de flexibilidad que se tolera en su diseño. La **tabla 19.1** ilustra algunas de esas diferencias.

TÉCNICAS DE INVESTIGACIÓN CUALITATIVA

Las tres técnicas de investigación cualitativa que se utilizan más a menudo en investigación en salud son: 1) la observación, 2) las entrevistas y 3) los grupos focales.

Cada método tiene utilidades diferentes: la observación es un método apropiado para recolectar datos sobre comportamientos que ocurren naturalmente en un contexto determinado. Las entrevistas son una estrategia metodológica ideal para recolectar información sobre perspectivas, actitudes y experiencias individuales e historias personales; particularmente, cuando se exploran temas muy sensibles. Los grupos focales son muy útiles para obtener información sobre las perspectivas y las normas culturales de un grupo.

Observación

Los métodos observacionales cualitativos se caracterizan por la observación directa, sistemática y detallada de la conducta y del habla, con el fin de describirla. Observar, en este contexto, significa examinar a un individuo, a un grupo de personas o un evento, con el fin de describirlo. La observación permite evaluar la discrepancia entre lo que las personas dicen y lo que en realidad hacen. Es un recurso metodológico que favorece el control de los sesgos inherentes a la tendencia de las personas a presentar una imagen favorable de sí mismas.

La observación, como técnica de recolección de datos cualitativos, puede clasificarse como: *observación participante* y ***observación no participante***; el grado de participación varía de acuerdo con las preguntas de investigación y con el contexto en el que se realiza el estudio.

En la ***observación participante***, el observador se convierte en "uno más" en la situación que está investigando. El investigador participa, y penetra así en la experiencia de otras personas y en los eventos que se estudian. Este abordaje metodológico implica un proceso difícil, formar parte de la situación sin perder la perspectiva investigativa que guía su presencia en este contexto[2].

Tabla 19.1 Diferencias entre los abordajes cualitativo y cuantitativo en investigación

	Abordaje cuantitativo	Abordaje cualitativo
Estructura	**El foco es conciso y preciso.**	**El foco es complejo y amplio.**
	Razonamiento lógico-deductivo.	Razonamiento dialéctico-inductivo.
Objetivo	Verificar hipótesis.	Explorar e interpretar fenómenos subjetivos.
	Cuantificar variaciones en los datos.	Describir variaciones en los datos.
	Predecir relaciones causales.	Describir y explicar relaciones.
	Describir las características de una población.	**Describir experiencias individuales o normas** de un grupo.
	Procura la generalizabilidad de los resultados.	**Procura la unicidad y la particularidad de los** hallazgos.
Flexibilidad en el diseño	La pregunta de investigación se define *a priori* y no se modifica en el curso del estudio.	**La pregunta de investigación se ajusta y se** refina a medida que se recolectan los datos.
	El diseño es relativamente estable y sigue un protocolo previamente establecido y sujeto a un conjunto de premisas estadísticas y otras condiciones.	El diseño es flexible y se ajusta a lo largo del estudio. Algunas preguntas pueden añadirse, suprimirse o modificarse conforme se avanza.
Muestreo	**El tamaño de muestra se define para alcanzar** un poder estadístico con significación clínica.	**Se busca una muestra rica en información.**
Formato de los datos	Numéricos, discretos o categóricos.	**Textuales (se obtienen mediante registros** audiovisuales y notas de campo).
Análisis	Las unidades de análisis son números o categorías.	**Las unidades de análisis son palabras, ideas y** comportamientos.
	El análisis se hace una vez la fase de recolección de datos ha concluido.	El análisis comienza desde cuando el investigador recolecta los datos; los organiza y los analiza al mismo tiempo.

Fuente: Elaboración de los autores.

La ***observación no participante*** ocurre cuando el investigador observa los eventos sin participar en las actividades de la situación o de las personas observadas. La ***observación no participante*** puede ser ***estructurada***, cuando los eventos que se van a observar han sido definidos antes del ingreso al "campo", y ***no estructurada*** cuando no se utiliza una guía predeterminada que guíe las observaciones, y se favorece el libre registro de los eventos en su contexto.

El proceso de recolección de datos, cuando se utiliza la observación como técnica de investigación, es inevitablemente selectivo, y depende, sobre todo, de la capacidad del investigador para actuar como un instrumento de investigación, con el fin de documentar la situación que observa. Para ello, es fundamental que las observaciones se registren y se analicen de manera sistemática durante su ocurrencia o poco tiempo después[2]. Habitualmente, el investigador utiliza un diario de campo, donde consigna de forma detallada sus observaciones de lo que sucede, sus propios sentimientos y sus respuestas ante los eventos, y sus cambios de perspectiva a lo largo del tiempo.

Los diarios de campo son muy detallados en las descripciones que recogen. En un principio, dichas notas pueden ser amplias, pero a medida que se avanza en el estudio, es posible que el investigador se enfoque en situaciones más específicas del contexto que observa. En la etapa final, y a medida que las hipótesis de trabajo se definen más claramente, el foco de observación puede ser mucho más restringido. La codificación y el análisis de las notas de campo es un proceso iterativo que implica la transformación de las observaciones en temas, categorías, conceptos, interpretaciones y tipologías, para darle un sentido a la situación, a los eventos y a las interacciones observados[2]. A menudo ocurre durante la misma fase de recolección de los datos, lo que permite evaluar, reformular y refinar las hipótesis de trabajo.

Un aspecto que no debe desatenderse cuando se utiliza la observación como técnica de investigación cualitativa es la manera como el investigador entra y sale de la situación de estudio. Sin importar el grado de participación del observador, el impacto que tiene su presencia en el contexto que se estudia acarrea, inevitablemente, una serie de reacciones en los sujetos. Es importante que el investigador sea claro y transparente respecto a los sujetos de observación, y que les explique la naturaleza de su presencia en ese contexto y sea consciente del impacto que puede generar y de las modificaciones que puede introducir.

También es crítico que cuente con el apoyo de un equipo de investigadores, que, si bien no se hallan expuestos directamente a la situación o a los eventos estudiados, le puedan proporcionar al investigador/observante/participante un punto de contraste que reduzca el riesgo de "volverse nativo" y sumergirse en la cultura del grupo, con la consecuente pérdida de su agenda de investigación.

Entrevistas

Si bien los clínicos entrevistan a sus pacientes durante el desarrollo de su trabajo asistencial, esa entrevista no puede ser equiparada a la que se usa como técnica de investigación cualitativa, y podría definirse como un uso general del término. La entrevista como técnica cualitativa es una modalidad bien establecida que busca comprender, a través del lenguaje y de la expresión no verbal de los individuos entrevistados, las perspectivas, las situaciones, los problemas, las opiniones y las experiencias en torno a un fenómeno de interés. La entrevista puede utilizarse de diferentes maneras en la investigación en salud: puede ser utilizada como técnica base para la recolección de datos en un estudio, o puede ser útil en estudios piloto para generar hipótesis y para obtener mayor información sobre un tema de interés o puede complementar la recolección de datos obtenidos mediante otras estrategias (cuantitativas o cualitativas).

Las entrevistas se clasifican en: ***estructuradas***, *semiestructuradas* y *no estructuradas/en profundidad*[3,4].

En las ***entrevistas estructuradas***, se administra una serie de preguntas previamente seleccionadas por el investigador, bajo la forma de un cuestionario. Los entrevistadores deben ser entrenados para hacer las preguntas de una forma estandarizada y de acuerdo con un formato preestablecido.

Las ***entrevistas semiestructuradas*** siguen un formato más laxo. El entrevistador cuenta con una guía de entrevista que sirve para tomar en cuenta todos los temas que va a desarrollar y que se consideraron relevantes. Las preguntas se plantean de forma más abierta, el entrevistador puede desviarse del formato y no es necesario mantener un orden estricto en su desarrollo, para poder profundizar en aspectos de particular interés que surjan durante la entrevista.

Las ***entrevistas no estructuradas o en profundidad*** utilizan un abordaje flexible para recoger información a partir de conversaciones.

Las entrevistas no estructuradas son de gran utilidad al comienzo de cualquier estudio cualitativo, pues ayudan a guiar decisiones sobre los aspectos centrales del problema o la situación que se quiere investigar.

Cuando se utilizan como herramienta principal en la recolección de datos cualitativos, las entrevistas no estructuradas permiten profundizar en los pensamientos, las emociones y las conductas de las personas acerca del tema de interés dentro del estudio.

Cuanto menos estructurada sea la entrevista, menos predeterminadas y menos estandarizadas son las preguntas. Existen varios tipos de preguntas que pueden plantearse en una entrevista[4]: las dirigidas a conductas o experiencias, a opiniones o valores, a sentimientos, a conocimientos, a experiencias personales; y las dirigidas a antecedentes personales, familiares o demográficos.

Las preguntas deben ser abiertas, naturales, sensibles y claras para el entrevistado. Una recomendación importante es comenzar con preguntas que sean fáciles de responder, para luego proceder con las más difíciles o que cubran aspectos más sensibles dentro de la situación de estudio. Cuando el formato de preguntas no es estandarizado, el orden puede variar de acuerdo con la evolución de la situación, o, incluso, pue-

de hacerse una nueva formulación, en la medida en que se haga evidente el tipo de vocabulario que las personas utilizan para describir los eventos o experiencias en cuestión, o para reaccionar a alguna situación o alguna información surgidas durante la entrevista

Cuando el estudio aborda situaciones muy dependientes del contexto cultural, se espera que el lenguaje utilizado por los individuos refleje su vivencia ante los eventos, por lo cual se recomienda que la entrevista sea sensible al lenguaje y a los conceptos utilizados por el entrevistado; incluso, conservándolos en el momento de plantear las preguntas.

Una variante significativa de las entrevistas cualitativas son las ***entrevistas conjuntas***. Son un método utilizado a menudo en estudios sobre relaciones maritales, vida en pareja o de familia, y últimamente su uso se ha extendido a explorar características generacionales de los entrevistados. Su diferencia básica respecto a la entrevista individual es que esta última es una reconstrucción individual de eventos y opiniones, en tanto que la entrevista conjunta reconstruye un panorama a partir de una díada o pareja de actores que participan de la misma situación, pero, posiblemente, con perspectivas distintas.

En ocasiones, los participantes en entrevistas conjuntas intentan adrede generar una imagen unificada de la realidad. Sin embargo, dichas entrevistas revelan diferentes clases de reacciones por parte de cada persona en torno a un fenómeno, y producen datos más completos y profundos acerca de las interacciones y de la naturaleza de las relaciones de poder entre los participantes.

La entrevista conjunta puede ser utilizada como parte de un diseño de investigación que use igualmente entrevistas individuales. Uno de los beneficios de combinar ambos tipos de entrevistas es que, en principio, se proporciona un sentido de igualdad a ambos participantes. No obstante, se recomienda que cuando se usen ambos esquemas de entrevista, se hagan primero las entrevistas individuales, y luego, la entrevista conjunta, para conocer algunos elementos cuya importancia, por su grado de sensibilidad, puede no haber sido mencionada en la entrevista conjunta, buscando evitar conflictos.

Otra variante de las entrevistas cualitativas son los ***diarios personales combinados con entrevistas***. Esta es una técnica confiable para obtener información sobre conductas, eventos y otros aspectos de la vida cotidiana de los sujetos que son difíciles de recordar y susceptibles de ser olvidados. Muchos estudios cualitativos han utilizado el método de diario de un día normal para conocer aspectos importantes sobre salud, enfermedad, conductas asociadas a la dieta y la nutrición, el uso de drogas y la conducta sexual.

Los diarios pueden ser de formato abierto, o, al contrario, muy estructurados, de manera que las actividades estén previamente categorizadas. Son útiles cuando otros métodos para evocar información han fracasado o cuando se trata de temas muy personales. Aunque son muy eficaces para recoger información, tienen la desventaja de depender mucho del grado de escolaridad del informante, así como de la aceptación y la adherencia del sujeto y de su motivación para registrar los eventos.

Grupos focales

La técnica de grupos focales tiene una larga trayectoria en el campo de las investigaciones de mercado y se ha convertido en una herramienta fundamental en el área de investigación en salud. Existen muchas definiciones en la literatura sobre lo que es un grupo focal, y, en términos generales, se identifica con atributos como la discusión organizada, la actividad colectiva, eventos sociales e interacciones[5].



Los grupos focales han sido definidos como "grupos de individuos seleccionados y ensamblados por los investigadores con el fin de discutir y comentar un asunto que es el objeto central de un estudio a partir de la experiencia individual de sus participantes"[6].

Parecería que los grupos focales son una forma grupal de entrevista, por lo que es importante resaltar las diferencias entre estas dos técnicas: en la entrevista grupal se busca entrevistar simultáneamente a un determinado grupo de personas prestando atención especial a las preguntas y las respuestas que surjan entre el investigador y los participantes. Los grupos focales se fundamentan en la interacción que se produce dentro del grupo al discutir los temas proporcionados por el moderador[5]. La interacción entre los participantes es la característica

distintiva de los grupos focales, pues pone en evidencia su manera de ver el mundo, el lenguaje que utilizan y los valores y sus creencias acerca de una situación. A partir de las preguntas planteadas entre unos y otros participantes y de sus interacciones, se posibilita reconsiderar su propia manera de apreciar ciertas experiencias, y se crea así un espacio para el intercambio y la transformación.

El objetivo principal de la investigación con grupos focales es obtener información acerca de las actitudes, los sentimientos, las creencias, las experiencias y las reacciones de los participantes acerca de un tema específico propuesto, o de una situación o una vivencia. Dichas características pueden ser parcialmente independientes de un grupo o de un contexto social determinado, pero es posible que sean más fáciles de revelar a través de la interacción social que el grupo focal estimula.

En comparación con las técnicas de observación, los grupos focales le permiten al investigador recolectar bastante información en un lapso más corto, sin que sea necesario esperar a que las cosas sucedan, pues el investigador sigue una guía de desarrollo del grupo focal.

Comparados con las entrevistas individuales (que buscan explorar actitudes, creencias y sentimientos de manera individual), los grupos focales permiten apreciar la complejidad y la multiplicidad de puntos de vista y de procesos emocionales dentro de un contexto grupal. Son particularmente útiles cuando existen diferencias de poder entre los participantes, cuando los asuntos de interés son el uso cotidiano del lenguaje y la cultura de un grupo particular, cuando se quiere explorar el grado de consenso en torno a un asunto determinado y, también, cuando se quieren tener diversas perspectivas acerca del mismo asunto.

Los grupos focales también les ofrecen a sus participantes la oportunidad de involucrarse en procesos de toma de decisiones, y la de ser considerados, en un momento dado, expertos, capaces de explorar soluciones frente a un problema particular.

Los grupos focales también tienen desventajas; algunas de ellas pueden superarse mediante una ejecución cuidadosamente planeada, pero otras son inevitables e inherentes a sus características. El investigador, por ejemplo, tiene menos control sobre los datos recolectados que el que puede tener mediante la observación o las entrevistas. Como el moderador debe permitir un alto grado de interacción entre los participantes, y, simultáneamente, debe intentar mantener a los participantes centrados en el asunto en discusión, el desarrollo de la actividad de un grupo focal no puede estar por completo predeterminado.

De igual manera, no puede asumirse que los participantes están expresando su propio punto de vista definitivo sobre cierto asunto, sino que están hablando en un contexto particular, dentro de una cultura específica, de modo que muchas veces se hace difícil para el investigador identificar claramente mensajes individuales en la discusión grupal. Los grupos focales son útiles pues para comprender la dinámica de un grupo como unidad en torno a un asunto de interés particular, pero no lo son para capturar la perspectiva individual de los participantes.

Desde un punto de vista práctico, los grupos focales pueden ser difíciles de ensamblar y requieren una planeación y una ejecución cuidadosas. Puede desestimularse la participación sistemática de cierto tipo de personas (bien sea por problemas de comunicación, o bien, porque no están dispuestas a compartir información personal con otros individuos), y en ese caso se recomienda utilizar otro tipo de técnicas cualitativas. Finalmente, los grupos focales no pueden ser del todo confidenciales o anónimos, pues se comparten material y reacciones con otras personas dentro del grupo, por lo cual es necesario controlar que la información de cualquiera de los participantes no sea utilizada luego en su contra, o fuera de contexto; en algunos casos, puede ser deseable para los participantes que tal información no se utilice nunca por fuera del grupo por parte de los participantes no investigadores.

La organización de un grupo focal requiere, de vez en cuando, una mayor planeación que otros métodos cualitativos. Es importante establecer ciertos parámetros para el desarrollo de los grupos focales y una guía de conducción para su desarrollo. El número de grupos y de reuniones puede variar: algunos estudios recurren a una reunión con varios grupos, o bien, a varios encuentros con el mismo grupo.

También reviste importancia considerar quiénes serán los participantes más adecuados para el grupo focal. Si el grupo es demasiado heterogéneo, las diferencias entre los participantes pueden tener un impacto considerable en

sus contribuciones, y el cual puede ser negativo. Alternativamente, si un grupo focal es muy homogéneo respecto a ciertas características específicas, puede reducirse la diversidad de opiniones y de experiencias proporcionadas, y se limitarán las opciones de interacción[7].

Los mencionados factores dependen mucho del asunto de interés dentro del estudio, de manera que no hay fórmulas sistemáticas para ensamblar un grupo focal. En cualquier caso, es indispensable que los participantes se sientan cómodos unos con otros. Los grupos focales donde los sujetos perciben a los demás como sus similares, con niveles de comprensión parecidos en torno a un asunto determinado, logran un desarrollo más exitoso. El fundamento de la selección está en lograr esa similitud benéfica sin llegar a la homogeneidad total.

El número de participantes que se recomienda por grupo es usualmente de seis a diez, aunque puede haber variaciones en las recomendaciones entre cuatro y quince personas[8].

La duración de una sesión de grupo focal varía entre una y dos horas, y sigue una secuencia caracterizada por:

- La presentación del asunto que se va a discutir.
- La fase de discusión propiamente dicha.
- La fase de cierre, en la que se hacen explícitos los elementos planteados por el grupo.

El papel del moderador en el proceso es crítico, ya que además de proporcionar explicaciones claras sobre el propósito del grupo focal y facilitar la interacción entre los participantes, debe estar atento a la dinámica de sus interacciones. Por ello, se requiere que el moderador tenga mucha capacidad de liderazgo y buenas habilidades en cuanto a relaciones interpersonales.

Es recomendable disponer dentro del grupo focal de un ***comoderador***, **quien, además de** encargarse de registrar la sesión, participe posteriormente en el análisis de los datos y ayude a identificar elementos que pudieron haber pasado inadvertidos para el moderador. Su participación en el grupo focal permite un mayor control sobre los contenidos considerados relevantes en el análisis de los datos, lo que es fundamental para darles validez a los hallazgos.

Por último, las consideraciones éticas planteadas en el uso de grupos focales son las mismas que se aplican para la mayoría de métodos de investigación social: los investigadores deben proporcionar toda la información acerca del propósito y del uso que se les dará a las contribuciones de los participantes; además, deben evitarse presiones a los participantes para que hablen, y debe garantizarse y mantenerse el anonimato de la información proporcionada por los informantes[9].

PASOS PARA LA EJECUCIÓN DE UN ESTUDIO CUALITATIVO

El proceso de un estudio cualitativo se compone de cuatro elementos:

- Establecimiento de un área de interés.
- Planteamiento de la pregunta de investigación.
- Recolección de datos.
- Procedimiento de análisis de la información obtenida.

El ***área de interés*** **se refiere al objeto de estudio** dentro del proyecto de investigación; por ejemplo, la interacción médico-paciente, la baja adherencia a un tratamiento o los patrones de prescripción de medicamentos por parte de los clínicos.

Por ***pregunta de investigación*** **se entiende la** pregunta que el estudio busca responder.

EJEMPLO

En el tema de la relación médico-paciente, se puede tener interés en cómo los médicos y los pacientes indagan sobre el estado de salud actual. En el caso de la baja adherencia a un tratamiento, la pregunta de estudio puede ser: *¿Cuáles son las actitudes de los pacientes acerca de la enfermedad que podrían explicar una baja adherencia al tratamiento?* **En cuanto a los patrones de prescripción** de medicamentos, la pregunta de interés podría ser: ***a partir de su experiencia personal, ¿qué*** *piensan los clínicos sobre los protocolos clínicos establecidos para prescribir medicamentos en un determinado grupo de pacientes?*

La ***recolección de datos*** **se refiere a las** técnicas utilizadas para obtener la información

necesaria para responder a la pregunta de investigación. Se trata de datos que surgen de la interacción entre el investigador y los participantes del estudio, del registro de sus opiniones, sus ideas y sus expectativas, o de las observaciones que quedan registradas.

EJEMPLO

En el tema de la interacción médico-paciente, los datos del estudio provienen de los registros de sus conversaciones, de la revisión de las notas del caso o del resultado de entrevistas a los médicos y a los pacientes acerca de dichas interacciones. En el caso de la baja adherencia a un tratamiento, puede recolectarse información sobre la evolución de los pacientes, indagar directamente con ellos sobre cómo perciben su enfermedad y el impacto que esta tiene en sus actividades cotidianas. En el caso de los patrones de prescripción de medicamentos, pueden revisarse las notas que los clínicos consignan en las historias clínicas o entrevistarles directamente.

El ***procedimiento para analizar los datos*** se refiere a la manera como se van a analizar los datos que se han recogido mediante las técnicas cualitativas. El análisis de datos cualitativos depende, principalmente, de su origen y su modo de recolección. Así como en la investigación cuantitativa es importante saber si los datos son ordinales o categóricos para hacer el análisis, en la investigación cualitativa también hay que saber si los datos se derivan de una entrevista directa o de observaciones de comportamientos. A partir de tal distinción, se puede seleccionar el tipo de análisis cualitativo que mejor se ajusta a cada caso.

VALIDEZ Y CONFIABILIDAD EN LA INVESTIGACIÓN CUALITATIVA

Antes de evaluar la validez de los datos, es importante considerar si la indicación de utilizar un abordaje cualitativo para el estudio es correcta, ya que en algunos casos puede ser más apropiado utilizar técnicas cuantitativas. También debe evaluarse si el análisis de los resultados es apropiado para el diseño, pues, más o menos a menudo, se ejecutan análisis que no corresponden. Por ejemplo, se analizan los datos cualitativos como si se tratara de un estudio cuantitativo, en cuyo caso, se resta profundidad al análisis de los aspectos indicativos de que el enfoque debía ser cualitativo, o se dificulta el análisis de datos que necesitan comprensión y explicación más que medición.



En la evaluación de la validez de un estudio cualitativo, debe evaluarse el papel que juega la subjetividad en su planeación, su ejecución y su análisis. Los estudios cualitativos son muy vulnerables a sesgos por parte del investigador.

Debido a que no hay análisis estadístico, es difícil verificar si el análisis de un proyecto con datos cualitativos es "correcto" o "incorrecto". Debido a la naturaleza propia de las técnicas cualitativas, es difícil determinar si se está haciendo bien el estudio. Dado el alto grado de flexibilidad, en este diseño se corre el riesgo de que el investigador poco disciplinado no profundice en los temas importantes del estudio. Como resultado de estos posibles escollos, la investigación cualitativa es controvertida, y el debate gira en torno a cómo asegurar su calidad.

Un buen estudio cualitativo emplea estrategias metodológicas variadas, y sus resultados son válidos cuando los hallazgos encontrados con diferentes técnicas de recolección de datos son congruentes. Por ejemplo, las respuestas que un individuo proporciona en una entrevista semiestructurada pueden compararse con material escrito (cartas, diarios personales u otros documentos) y con la manera como la persona responde a las contribuciones de otros miembros de un grupo de discusión. Esta validación cruzada de la información se conoce como *triangulación*.



La triangulación como estrategia permite reducir sustancialmente el riesgo de que los hallazgos de un estudio cualitativo sean producto de un método aislado de recolección de los datos, de una fuente única de información o de los sesgos del investigador[8,10].

Es una estrategia que aumenta la validez del estudio cualitativo y mejora la confianza en la autenticidad de sus datos y de su interpretación[11]. **En términos prácticos, significa** que los investigadores utilizan fuentes

independientes de información o de conjuntos de datos, diferentes tipos de análisis, a diferentes investigadores o diferentes perspectivas teóricas para estudiar un fenómeno particular. En la **tabla 19.2 se presentan diferentes tipos de** triangulación que se utilizan en la investigación cualitativa[8].

Las diferencias entre los resultados obtenidos con varias técnicas cualitativas (por ejemplo, entrevistas y grupos focales) no necesariamente ponen en duda la validez de un estudio cualitativo. En algunas circunstancias, dichas diferencias agregan textura a los hallazgos.

EJEMPLO

Una persona expresa una opinión extrema en una entrevista individual, pero cambia a un punto de vista menos radical cuando participa en un grupo focal. Esta variación invita a interpretaciones acerca de la manera como el individuo se relaciona con un grupo de pares.

La investigación cualitativa ha sido ampliamente criticada por su falta de ***generalizabilidad.*** Para muchos, los datos obtenidos mediante técnicas cualitativas no pasan de ser un conjunto de impresiones y anécdotas personales, sujetas a sesgos del investigador y carentes de reproducibilidad. El supuesto subyacente a dichas críticas es que los abordajes cualitativos y cuantitativos son fundamentalmente disímiles en cuanto a los medios disponibles para asegurar la validez y la confiabilidad de sus hallazgos. Sin embargo, tal distinción es más de grado. También puede ocurrir que en el análisis de datos cuantitativos se generen representaciones estadísticas de un fenómeno bajo estudio que, como sucede en el trabajo cualitativo, dependerán del juicio y las habilidades del investigador y de la conveniencia de los datos para responder la pregunta en cuestión.

En la investigación cualitativa también hay maneras de garantizar, hasta cierto punto, la confiabilidad de los procesos de recolección, análisis e interpretación de los datos. Un registro claro de los datos y de las técnicas utilizadas, así como una cuidadosa documentación del proceso de análisis, permitirá que otro investigador llegue a conclusiones similares y produzca una explicación plausible y coherente del fenómeno estudiado. El desarrollo de taxonomías y de sistemas de codificación permite proteger la integridad de los estudios cualitativos a lo largo de todas sus etapas.

ESTRATEGIAS DE MUESTREO EN LA INVESTIGACIÓN CUALITATIVA

La lógica del muestreo en estudios cualitativos es diferente de la utilizada en investigaciones cuantitativas.

Tabla 19.2 Tipos de triangulación

Triangulación de las fuentes de datos	Triangulación de los investigadores
• Se compara y se verifica la consistencia de la información en diferentes momentos, a partir de diferentes técnicas y con diferentes fuentes. • Se compara la información obtenida mediante entrevistas, observaciones y documentos escritos. • Se compara lo que dicen las personas en público con lo que dicen en privado. • Se verifica la consistencia de lo que las personas dicen o hacen a lo largo del tiempo. • Se comparan las perspectivas de los participantes respecto a diferentes puntos de vista.	Se recurre a más de una persona para recolectar y analizar los datos: • Varias personas hacen las observaciones del mismo fenómeno para protegerse de sesgos de observación. • Participan diferentes entrevistadores: los datos obtenidos no son un artificio de un estilo particular de entrevista o de una percepción individual. • Dos o más investigadores analizan independientemente los mismos datos cualitativos, y luego comparan sus hallazgos. • Se permite que los participantes revisen los hallazgos.
Triangulación de los métodos	**Triangulación de la teoría**
• Se obtienen los datos utilizando diferentes técnicas (entrevistas en profundidad, grupos focales y observación).	• Se utilizan diferentes abordajes para organizar y enfocar el análisis. • Se utilizan múltiples perspectivas o teorías para interpretar los datos.

Fuente: Elaboración de los autores.

El punto crucial para el muestreo es la identificación de grupos específicos de personas que tengan las características requeridas para el estudio, de acuerdo con unos criterios de inclusión o de exclusión, y que sean una fuente rica de información[12].

Dicho abordaje de muestreo en la investigación cualitativa le permite al investigador incluir de forma deliberada una amplia variedad de informantes para acceder a fuentes importantes de conocimiento sobre el tema en estudio.

El muestreo cualitativo tiene varias características que lo definen:

Su diseño, aunque preconcebido, es flexible y se ajusta en la medida en que el estudio progresa.

Las unidades de muestreo se seleccionan secuencialmente. El que viene después depende del que venía primero.

La muestra se ajusta continuamente, de acuerdo con los desarrollos de la teoría y de los resultados que se van obteniendo. Se hace una selección inicial de informantes, y una vez recolectados, codificados y analizados sus datos, se produce una explicación teórica preliminar, antes de decidirse si se requieren más datos y a partir de qué tipo de muestra se requieren. La relación entre el muestreo y la explicación es iterativa y teóricamente orientada[13].

La selección continúa hasta el punto de redundancia.

La muestra incluye la búsqueda de casos negativos con el fin de darle mayor fuerza y amplitud al desarrollo de la teoría.

Existen diversas opciones para hacer muestreos en estudios cualitativos[8,12,14], **y que se** presentan en la **tabla 19.3.**

ANÁLISIS EN ESTUDIOS CUALITATIVOS

Una vez se han recolectado los datos cualitativos, el investigador debe seguir una serie de pasos para su análisis:

Búsqueda de vacíos de información.

Transcripción del material (de manera literal, incluyendo pausas, información de tipo no verbal, impresiones del investigador, etc.).

Registro de los datos en un procesador de palabras.

Limpieza de los datos (cambio de los nombres de los participantes para garantizar la confidencialidad de la información), y revisión de errores tipográficos o de transcripción.

Codificación de los datos.

Identificación de temas y subtemas.

Desarrollo de categorías provisionales.

Descripción del fenómeno de estudio.

Exploración de relaciones entre las categorías.

Refinamiento de los temas y de las categorías.

Desarrollo de la teoría e incorporación de conocimiento previo.

Evaluación y validación de modelos de explicación mediante la triangulación de datos, de fuentes y de abordaje analítico.

Redacción del reporte (incluyendo viñetas a partir de los datos originales, en caso de ser apropiado).

El abordaje analítico en los estudios cualitativos es de tipo inductivo, de modo que los datos orientan la interpretación y la teoría emergente.

El análisis de datos cualitativos conlleva un proceso iterativo de investigación que implica volver a los datos para ajustar y delimitar su campo[15]. **A veces, las preguntas de mayor** interés se hacen evidentes al final del proceso de investigación; entonces suele decirse que los datos enseñan al investigador y orientan el proceso de análisis y explicación, de una manera dinámica y flexible[1].

Las técnicas analíticas para datos cualitativos pueden dividirse en dos grandes categorías:

Técnicas de análisis basadas en teorías lingüísticas.

Técnicas basadas en la *tradición sociológica*[16].

Las técnicas de análisis basadas en la lingüística abordan el texto como un objeto de análisis ***per se*, e incluyen métodos como el análisis narrativo,** el análisis del discurso y el análisis lingüístico formal. Las técnicas de análisis basadas en la sociología abordan el texto como si este fuera una ventana de acceso a la experiencia humana,

Tabla 19.3 Opciones para hacer el muestreo en estudios cualitativos

Tipos de muestreo	Utilidad
Muestreo de casos extremos.	Caracterizar manifestaciones inusuales de un fenómeno de interés.
Muestreo por intensidad.	Caracterizar casos que manifiestan intensamente el fenómeno de interés.
Muestreo según variaciones máximas.	Permite obtener información a partir de un amplio rango de variación del fenómeno de interés.
Muestreo homogéneo.	Reduce la variación y simplifica el análisis.
Muestreo de casos típicos.	Ilustra lo que es típico, normal o promedio.
Muestreo por conveniencia, estratificado.	Ilustra características de subgrupos particulares de interés; facilita comparaciones.
Muestreo de casos críticos.	Permite una generalización lógica y la máxima aplicación de la información a otros casos.
Muestreo de "bola de nieve".	Identifica casos de interés a partir de personas que conocen a otras personas quienes, a su vez, conocen a personas que pueden ser buenos casos para el estudio.
Muestreo según criterio.	Captura todos los casos que cumplen algún criterio; asegura la calidad de la información.
Muestreo basado en constructos operacionales o en teorías.	Busca manifestaciones de un constructo teórico de interés.
Casos confirmatorios o desconfirmatorios.	Elabora y profundiza análisis iniciales. Sirve para identificar excepciones.
Muestreo oportunista.	Toma ventaja de los imprevistos.
Muestreo de casos políticamente importantes.	Atrae atención sobre el estudio, o bien la desvía mediante la eliminación de casos políticamente sensibles.
Muestreo por conveniencia.	Ahorra tiempo y dinero.
Muestreo combinatorio.	Utiliza diferentes estrategias de muestreo para triangular.

Fuente: Elaboración de los autores.

y dentro de dicha categoría se incluyen, a su vez, tres subcategorías:

El análisis de palabras o frases obtenidas mediante técnicas estructuradas de recolección de datos. Para ello, se pueden utilizar varias estrategias analíticas, como el análisis de componentes, la construcción de taxonomías o de mapas mentales. Por ejemplo, una estrategia consiste en contar el número de veces que una palabra en particular ("deprimido") o un concepto ("depresión") ocurren en una narración.

El análisis de "textos libres", como narrativas, discursos y respuestas a preguntas abiertas durante entrevistas, y para los que se puede utilizar análisis de redes semánticas o análisis de contenido. Por ejemplo, en un análisis de contenido, todas las unidades de datos (frases o párrafos) en los que un individuo hace referencia a la tristeza recibirían un código particular y se examinarían con más detalle: ¿el individuo habla de depresión relacionada con experiencias particulares? ¿Qué términos utiliza para describir la depresión? ¿Existen diferencias entre los participantes en la manera como se refieren a la depresión?

El análisis basado en la teoría fundamentada[17]. Este también utiliza sistemas de codificación, pero los códigos se contrastan con teorías ya existentes, para interpretar los datos. Por ejemplo, al analizar la información que los participantes proporcionan acerca de la depresión, el investigador encuentra que la depresión estaría casi sistemáticamente relacionada con la pérdida de un ser querido. En ese caso, sería necesario contrastar dicha teoría emergente con otras teorías planteadas en la literatura sobre el tema. En algunos casos puede ser necesario tratar de identificar

casos que contradigan la teoría, y, a la luz de ello, modificarla. Este procedimiento se conoce como ***inducción analítica***, y es utilizado para construir y evaluar teorías emergentes a partir de los datos[18].

Naturalmente, siempre es posible combinar diferentes abordajes analíticos dentro de un proyecto y utilizar más de una estrategia para examinar el mismo cuerpo de datos. La sugerencia que puede darse aquí es tratar de actuar de la manera más simple posible si no se tiene mucha experiencia; se recomienda intentar algunos abordajes analíticos comunes, que ya hayan sido empleados, antes de intentar métodos más sofisticados de análisis.

ANÁLISIS SECUNDARIO DE DATOS CUALITATIVOS

Las oportunidades limitadas (y, en ocasiones, los costos) para el desarrollo de estudios cualitativos primarios han estimulado a los investigadores a considerar la opción del uso de datos previamente recolectados.



El análisis secundario de datos cualitativos se hace con datos ya existentes, recolectados para un estudio anterior, y busca responder a una pregunta de interés distinta de la del trabajo original o darle un matiz diferente o mayor profundidad a la pregunta que originó el estudio.

El advenimiento de diferentes programas de computador para la asistencia en la codificación, la recuperación y el análisis de datos cualitativos es otro desarrollo que ha facilitado el acceso a conjuntos de datos cualitativos susceptibles de usarse con propósitos de análisis secundarios.

Existen diferentes estrategias de análisis secundario de datos cualitativos:

- El ***análisis adicional en profundidad***, el cual es un análisis más intensivo y profundo de un hallazgo particular o de un aspecto que fue desarrollado inicialmente como parte de un trabajo cualitativo primario.
- El ***análisis adicional de un subconjunto de datos***; o sea, un análisis sobre un subgrupo particular de la muestra del estudio original que comparta ciertas características que permitan hacer análisis adicionales.
- La ***aplicación de una nueva perspectiva o abordaje conceptual***; o sea, el análisis retrospectivo de una parte o de todo un conjunto de datos con una perspectiva diferente, para examinar conceptos que no fueron centrales en el estudio original.

Existen, por otra parte, algunas consideraciones de carácter metodológico y ético que deben ser tenidas en cuenta cuando se hacen análisis secundarios de datos cualitativos[19]. El primer aspecto que se debe considerar es el cuestionamiento que se hace a esta modalidad de análisis es sobre su factibilidad, pues, por su naturaleza, la investigación cualitativa implica una relación intersubjetiva entre el investigador y los participantes. Como respuesta a dicha crítica, podría argumentarse que aun cuando los datos primarios fueron recogidos a partir de entrevistas o de observación, por lo general, hay más de un investigador involucrado en esa clase de estudios, que participa en la contextualización y en la interpretación de los datos a pesar de no haberse hallado presente en el trabajo de campo.

Un segundo aspecto es el referente al problema de cuándo termina un análisis primario y cuándo comienza un análisis secundario de datos cualitativos. La investigación cualitativa es un proceso iterativo, usualmente basado en una teoría en particular, que requiere que las preguntas sean sometidas a un proceso de formulación y ajuste a lo largo del estudio[20]. Para los investigadores primarios, quienes reutilizan los mismos datos, puede ser difícil determinar si la investigación forma parte del estudio original o si es lo suficientemente nueva y diferente como para calificarla de análisis secundario.

También se han planteado algunas consideraciones de carácter ético y práctico sobre el análisis secundario de datos cualitativos, las cuales deben ser tenidas en cuenta[20,21]:

- La compatibilidad de los datos con el análisis secundario: ¿son los datos susceptibles de ser sometidos a un análisis secundario? Ello depende del ajuste entre el objetivo del análisis y la naturaleza y la calidad de los datos originales.
- ¿Cuál es la posición del investigador que analiza los datos secundarios? ¿Formó parte del estudio original? Dicha información influirá en la decisión de llevar a cabo o no este tipo de análisis y en la del tipo de procedimientos

que serán utilizados. Para un investigador que participó en el estudio original, la tarea de recurrir a las fuentes originales de datos será mucho más fácil. Cuando esa no es la situación, el investigador debe hacer consultas con el investigador primario para evaluar la calidad del trabajo original y contextualizar el material.

- Presentación del análisis original y del análisis secundario: es importante que el diseño del estudio, los métodos y los aspectos involucrados sean reportados en su totalidad. Eso incluye un delineamiento del estudio original y de los procedimientos de recolección de datos empleados, junto con una descripción del proceso de categorización y resumen de los datos usados en el análisis secundario.
- Cuando el objeto de estudio involucra datos sensibles, el consentimiento informado no puede presumirse. Aunque no siempre es posible obtener un consentimiento informado adicional para el análisis secundario, es importante contar con un juicio profesional (idealmente, el de un comité de investigación y ética institucional) acerca de si el uso de los datos para un análisis secundario viola o no el contrato entre los sujetos y los investigadores primarios[19]. Ese interés creciente por reutilizar datos cualitativos hace cada vez más necesario que los investigadores obtengan un consentimiento informado por parte de los participantes, para cubrir la posibilidad de un análisis secundario, lo cual es consistente con las recomendaciones éticas para la investigación en salud. Dicho consentimiento debe ser muy claro y tener explícitas las opciones de uso de la información en el futuro, y debe haber sido aprobado por un comité institucional de investigación y ética.

PROGRAMAS INFORMÁTICOS PARA ESTUDIOS CUALITATIVOS

Los programas informáticos para el análisis de datos cualitativos pueden clasificarse según el tipo de procesos analíticos que facilitan: recuperación de texto, almacenamiento de texto, codificación, construcción de teorías a partir de los códigos, y construcción de redes conceptuales[1]. En general, dichos programas buscan facilitar algunas de las tareas mecánicas de búsqueda y selección de datos y de identificación de textos para su codificación. Decidir cuál programa utilizar dependerá, principalmente, del tipo de análisis que se planea, pero también, de los recursos disponibles y de la cantidad de datos que se espera tener.



Los programas más populares que se utilizan en estudios cualitativos en salud son AtlasTi, NUD*IST y NVivo[22].

Es importante aclarar que los mencionados programas informáticos no pueden crear por sí mismos un esquema de codificación ni determinar cuál es el sentido de asignar ciertas categorías a un segmento particular de datos[23]. Sin embargo, su uso tiene varias ventajas: rapidez y flexibilidad para codificar la información, el hecho de que permiten más tiempo para las tareas de interpretación y que proporcionan una forma más compleja de examinar las relaciones entre los datos y de desarrollar una estructura formal para su análisis[24].

Una de las críticas más grandes que se le hacen a este tipo de programas[23,25] es que, según se considera, distancian a los investigadores de los datos. Otras limitaciones incluyen la pérdida de información de tipo no verbal y la directamente derivada de la relación entre el investigador y el participante.

Independientemente de las controversias generadas por el uso de programas informáticos en investigación cualitativa, es importante reconocer que no es posible analizar los datos con estas herramientas sin antes haberse familiarizado con la información a partir de su fuente original, que es la transcripción de los datos. Si bien los programas informáticos ayudan en la administración de la información y en el archivo de los datos, únicamente proporcionan asistencia en el análisis y en las elaboraciones conceptuales derivadas de este, pero jamás podrán sustituir la actividad creativa y personal del investigador.

REPORTE DE UN ESTUDIO CUALITATIVO

A continuación se proporcionan algunas pautas que sirven de punto de partida a los investigadores interesados en hacer estudios cualitativos:

A. Introducción

- Introduzca su pregunta de estudio. Describa qué es lo que se quiere estudiar y cómo nació su interés por el tema.
- Mencione por qué es necesario el estudio. Cite literatura relevante que sustente la necesidad de investigación en esta área, o los vacíos de conocimiento o la existencia de percepciones o de concepciones que pueden ser erradas. En sus propias palabras, describa por qué piensa que el estudio planteado puede ser útil.
- Describa cuál es el objetivo de su investigación (ampliar la base del conocimiento, generar hipótesis para un estudio cuantitativo, un desarrollo basado en una teoría…).

B. Paradigma de investigación

Esta sección debe incluirse en su propuesta cuando se asume que los lectores no están familiarizados con el paradigma de investigación cualitativa. No es necesaria en los contextos donde la investigación cualitativa es una forma aceptada de investigación.

- Utilice un lenguaje específico para nombrar y describir su paradigma de investigación. El término *paradigma* se utiliza aquí para representar las bases epistemológicas y conceptuales de su investigación cualitativa (naturalista, pospositivista, etc.).
- Explique el tipo de relación que el investigador tiene con los informantes (observador, observador participante, colaborador, etc.).

C. Método de investigación

- Identifique y describa de manera general el método que ha seleccionado (estudio de campo, estudio de casos) y sus procedimientos (entrevistas, grupos focales, observación).
- Describa en detalle qué pretende hacer desde cuando comienza el estudio.
- Explique cómo seleccionará a sus informantes y cómo ingresará dentro del contexto que será investigado (en caso de ser relevante).
- Describa los procedimientos mediante los cuales protegerá los derechos de sus informantes, la confidencialidad, la autonomía, el uso futuro de la información y los destinos que se les darán a los resultados (procedimiento de consentimiento informado).
- Describa la clase de relación que establecerá con sus informantes. ¿El investigador planea ser neutral, colaborador u objetivo?
- Describa la clase de datos que recolectará (notas de campo, grabación en audio, grabación en video, transcripción de conversaciones, examen de documentos preexistentes, etc.).
- Describa los procedimientos de recolección de los datos. Si va a utilizar entrevistas, presente sus preguntas. Describa los instrumentos que utilizará. Describa los procedimientos que le permitirán mantener un registro del proceso de investigación.
- Describa los procedimientos que piensa utilizar para el análisis (codificación, clasificación).
- Describa cómo organizará y presentará sus datos, sus interpretaciones y sus conclusiones.
- Describa cómo considerará y protegerá la confiabilidad y la validez. ¿Utilizará métodos y procedimientos sistemáticos? ¿Triangulación? ¿Auditoría? ¿Revisión de pares?

D. Sesgos preliminares, hipótesis y supuestos

- Describa cómo ha influido la revisión de la literatura en la manera como se ha desarrollado la investigación.
- Discuta de qué manera su experiencia previa con el tema de estudio ha influido en la forma como ha planteado la investigación. Puede resumir experiencias personales y profesionales relevantes, si no las ha mencionado en la introducción.
- Exponga los hallazgos que espera encontrar, sus hipótesis y sus supuestos de estudio.
- Discuta las limitaciones de su estudio en el contexto de las limitaciones de estudios similares en el área.

Recientemente, se han desarrollado criterios consolidados para el reporte de estudios cualitativos[26] que hacen parte de la Red Equator. Se invita al lector a conocer su contenido, para familiarizarse con buenas prácticas de publicación de este tipo de estudios.

APRECIACIÓN CRÍTICA DE ESTUDIOS CUALITATIVOS

Hay diferentes criterios para evaluar la calidad de un estudio cualitativo. Existe mucha controversia acerca de cuál es el mejor abordaje. Algunos autores plantean que no es posible determinar la calidad de un estudio cualitativo siguiendo criterios preestablecidos, mientras que otros plantean la necesidad de contar con abordajes más rigurosos para apreciar la validez y la confiabilidad de tal tipo de estudios[27,28].

Hay publicados en la literatura científica más de 100 instrumentos para evaluar la calidad de estudios cualitativos[27,29,30]. Esta diversidad de criterios refleja la falta de consenso acerca de qué se considera una investigación cualitativa de calidad, pero también es consecuencia de la necesidad de tener distintos tipos de instrumentos de evaluación de acuerdo con los métodos cualitativos que se utilicen[31].

Mays y Pope[32] plantean una serie de guías generales para evaluar la calidad de un estudio cualitativo:

1. **¿El investigador explicó la estructura teórica** y los métodos utilizados en cada fase del estudio?
2. **¿El contexto fue claramente descrito?**
3. **¿La estrategia de muestreo fue claramente** descrita y justificada?
4. **¿La estrategia de muestreo fue teóricamente** comprehensiva como para garantizar la generalización del análisis conceptual?
5. **¿Cómo se llevó a cabo el trabajo de campo?** ¿Fue descrito en detalle?
6. **¿La evidencia (notas de campo, transcripciones** de entrevistas, registros, etc.) puede ser examinada independientemente por otros? Si es relevante, ¿el proceso de transcripción puede ser inspeccionado de manera independiente?
7. **¿Los procedimientos para el análisis de** datos fueron claramente descritos y teóricamente justificados? ¿Cómo fueron identificados los conceptos y los temas a partir de los datos?
8. **¿El análisis fue repetido por más de un in**vestigador para garantizar la confiabilidad?
9. **¿El investigador proporcionó evidencia** para buscar casos que podrían contradecir o modificar el análisis?

También se recomienda la lista elaborada por Anderson[33] para ayudar a autores y a revisores de estudios cualitativos.

CONSIDERACIONES FINALES

Los métodos cuantitativos y cualitativos en investigaciones en salud son abordajes que, si bien se basan en paradigmas filosóficos diferentes, no necesariamente son excluyentes. Su relación no es jerárquica, sino complementaria; es decir, al utilizarse diseños mixtos (cualicuantitativos o cuanticualitativos) puede trabajarse en torno a complejos temas de investigación, para lograr un panorama más amplio y profundo de un problema de salud en particular.

La decisión de utilizar o no un diseño cualitativo o cuantitativo tiene un componente filosófico, que depende de la visión del mundo a la cual se suscribe el investigador. Pero depende también, y fuertemente, del tipo de pregunta investigativa planteada, que, a su vez, depende del objeto de la investigación y del tipo de datos que se van a recolectar para el estudio.

Es posible, y muy deseable para los intereses de un mayor pluralismo metodológico, explorar las suposiciones que subyacen a todas las investigaciones en el contexto de su utilidad, para comprender problemas particulares de investigación. Necesitamos estar abiertos y reconocer nuestras posiciones epistemológicas y las de otros investigadores, de manera que los usuarios de investigaciones tengan la oportunidad de participar completamente de sus hallazgos.



La selección de la metodología debe hacerse de forma que se ajuste al objeto de investigación y a la pregunta del proyecto, y debe evitarse la polarización de posiciones. Cada pregunta tiene su enfoque más apropiado, y es responsabilidad del investigador descubrirlo y utilizarlo.

Referencias

1. Watkins DC. Qualitative research: the importance of conducting research that doesn't "count". Health Promot Pract. 2012;13:153-8.
2. Bogdewic S. Participant observation. En: Crabtree B, Miller W, editores. Doing qualitative research. Thousand Oaks CA: Sage; 1999. p. 47-69.
3. Trost J. Perspectives on qualitative interviews. Symbolic Interaction. 2014;37:432-3.
4. Pope C, Mays N, Britten N. Qualitative interviews. En: Pope C, Mays N, editores. Qualitative research in health care. 3rd ed. Oxford: Wiley-Blackwell Publishing Ltd; 2006.
5. Then KL, Rankin JA, Ali E. Focus group research: what is it and how can it be used? Can J Cardiovasc Nurs. 2014;24:16-22.
6. Powell RA, Single HM. Focus groups. Int J Qual Health Care. 1996;8:499-504.
7. Freeman T. 'Best practice' in focus group research: making sense of different views. J Adv Nurs. 2006;56:491-7.
8. Patton MK. Qualitative research and evaluation methods. 3rd ed. Thousand Oaks, CA: Sage Publications; 2002.
9. Shaw IF. Ethics in qualitative research and evaluation. J Soc Work. 2003;3:9-29.
10. Reeves S, Kuper A, Hodges BD. Qualitative research methodologies: ethnography. BMJ. 2008;337:a1020.
11. Stoto M, Nelson C, Klaiman T. Getting from What to Why: using qualitative research to conduct public health systems research. Washington, DC: AcademyHealth; 2013.
12. Robinson OC. Sampling in interview-based qualitative research: a theoretical and practical guide. Qual Res Psychol. 2014;11:25-41.
13. Glaser BG, Strauss AL. The discovery of grounded theory: strategies for qualitative research. Hawthorne, NY: Aldine de Gruyter; 1967.
14. Morse J. Strategies for sampling. En: Morse J, editor. Qualitative nursing research: a contemporary dialogue. Londres: Sage; 1991.
15. Bernard HR, Ryan GW. Analyzing qualitative research: systematic approaches. Sage Publications: Thousand Oaks, CA; 2010.
16. Ryan GW, Bernard HR. Data management and analysis methods. En: Handbook of qualitative research, 2nd ed., Denzin N, Lincoln Y, editores. Thousand Oaks, CA: Sage Publications; 2000. p. 769-802.
17. Bryant A, Charmaz KC. Grounded theory in historical perspective: an epistemological account. En: Charmaz KC, Bryant A, editores. The SAGE handbook of grounded theory (paperback edition). Londres: SAGE Publications Ltd.; 2010. p. 31-57.
18. Charmaz K. Grounded theory. En Smith JA, editor. Qualitative psychology: a practical guide to research methods. London: SAGE Publications Ltd.; 2003. p 81-110.
19. Coltart C, Henwood K, Shirani F. Qualitative secondary analysis in austere times: ethical, professional and methodological considerations. Forum Qual Soc Res. 2013;14:18.
20. Corti L, Thompson P. Secondary analysis of archived data. En: Seale C, Gobo G, Gubrium JF, et al., editores. Qualitative research practice: concise paperback edition. Londres: SAGE Publications Ltd.; 2006. p. 297-313.
21. Heaton J. Reworking qualitative data. Londres: Sage Publications; 2004.
22. Lacey A, Luff D. Trent focus for research and development in primary health care: an introduction to qualitative analysis. Beterborough, ON: Trent Focus; 2001.
23. Cope DG. Computer-assisted qualitative data analysis software. Oncol Nurs Forum. 2014;41:322-3.
24. Seror J. Computers and qualitative data analysis: paper, pens, and highlighters vs. screen, mouse, and keyboard. **TESOL Quarterly. 2005;39:321-8.**
25. Jones M. Using software to analyse qualitative data. J Qual Res. 2007;1:64-76.
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19:349-57.
27. Dixon-Woods M, Shaw RL, Agarwal S, et al. The problem of appraising qualitative research. Qual Saf Health Care. 2004;13:223-5.
28. Walsh D, Downe S. Appraising the quality of qualitative research. Midwifery. 2006;22:108-19.
29. Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. Ann Fam Med. 2008;6:331-9.
30. Cote L, Turgeon J. Appraising qualitative research articles in medicine and medical education. Med Teach. 2005;27:71-5.
31. Kuper A, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. BMJ. 2008;337:a288.
32. Mays N, Pope C. Quality in qualitative research. En: Pope C, Mays N, editores. Qualitative research in health care. Oxford: Blackwell Publishing/BMJ Books; 2006. p. 82-101.
33. Anderson C. Presenting and evaluating qualitative research. Am J Pharm Educ. 2010;74:141.



SECCIÓN

4

Diseños de investigación clínica. Estudios secundarios o integrativos

CAPÍTULO 20

María Ximena Rojas Reyes, Juan Manuel Lozano León

CAPÍTULO 21

Hoover Quitián Reyes

CAPÍTULO 22

Óscar Mauricio Muñoz Velandia,
Ana María de la Hoz Bradford

CAPÍTULO

20

Revisiones sistemáticas y metaanálisis de la literatura

María Ximena Rojas Reyes • Juan Manuel Lozano León

INTRODUCCIÓN

Los avances de la ciencia deben ayudar a dar sentido a lo que ya sabemos y hemos observado[1]. **Sin embargo, en el área de la** salud no se puede desconocer que el acelerado avance de la investigación biomédica lleva a que la información sobre un tema tenga un crecimiento exponencial y resulte casi imposible para el clínico estar al tanto de todos y cada uno de los avances en su área de interés. Esto, sin lugar a dudas, retarda la aplicación del nuevo conocimiento al cuidado de los pacientes. La cantidad de información, la diversidad de fuentes de consulta y la variabilidad en la calidad de los estudios que se publican son algunas de las principales dificultades que encuentran los profesionales de salud para obtener y resumir eficientemente la información que se publica y que requieren para la toma de decisiones.

Para basar en evidencia la práctica clínica, los profesionales de la salud (médicos, odontólogos, enfermeros) necesitan no solo que haya información relevante y apropiada en su área de pericia; también necesitan habilidades para buscar la literatura y conocimientos sobre metodología de investigación que les permitan hacer apreciaciones críticas y síntesis de los estudios publicados, para poder aplicar los resultados en el cuidado de sus pacientes. Si bien el número de profesionales con entrenamiento para completar dichos procesos va en aumento, en la vida diaria no siempre es factible hacerlos, por las limitaciones en tiempo para responder a los cuestionamientos que surgen día a día en la práctica clínica.

Las revisiones sistemáticas de la literatura (**RSL**) **son una aproximación metodológica** para resumir y sintetizar los resultados de la investigación sobre una cuestión particular[2]. Se trata de un diseño con estrategias rigurosas que buscan minimizar los sesgos en los métodos con el fin de buscar estudios publicados, para su selección y su apreciación, así como para el resumen de sus resultados.



Las revisiones sistemáticas de la literatura (RSL) son una aproximación metodológica para resumir y sintetizar los resultados de la investigación sobre una cuestión particular.

Estas características han convertido a las RSL en un pilar fundamental para la toma de decisiones en salud basadas en evidencia. No solo son una de las principales fuentes de consulta para la práctica clínica basada en evidencia, sino que son la base para desarrollar recomendaciones en guías de práctica clínica, realizar evaluaciones de tecnologías en salud, evaluaciones económicas para definir el impacto económico de las intervenciones en salud y producir información para los tomadores de decisiones y los formuladores de políticas públicas[3].

Este capítulo presenta una introducción a los principios metodológicos que guían la conducción de RSL y su aplicación en los diferentes escenarios de toma de decisiones en salud. Se basa en una revisión rigurosa de los textos publicados por autores destacados en el tema[4,5], y del manual de la que se considera la organización con mayor experiencia en la preparación, la edición y la publicación de RSL: la Colaboración Cochrane[6]. Está dirigido a cualquier profesional de salud que quiera iniciarse en el desarrollo de RSL y a quienes busquen adquirir una actitud crítica hacia las RSL.

DEFINICIÓN

La RSL es un diseño integrativo cuyo fin es localizar y sintetizar, de forma exhaustiva, la investigación que se ha hecho sobre una cuestión determinada[4]. El adjetivo "sistemático" se refiere a que la revisión se prepara a partir de una pregunta claramente formulada y se conduce siguiendo una metodología explícita y reproducible (idealmente, consignada en un protocolo) para la identificación, la selección y la evaluación de estudios pertinentes, así como para la síntesis de los datos obtenidos de estos. Dicha metodología permite minimizar los errores sistemáticos (sesgos) y aleatorios a los que las revisiones de la literatura son susceptibles[7].

La aproximación sistemática y explícita es lo que distingue a las RSL de otras revisiones (revisiones narrativas tradicionales) por sobre un tema clínico publicado en la literatura biomédica[8]. En las revisiones narrativas tradicionales no suele haber una pregunta clara para responder, ni se describen los métodos empleados por los autores para localizar, seleccionar, evaluar ni resumir los estudios revisados. Las RSL presentan grandes ventajas sobre las revisiones narrativas en el momento de evaluar la evidencia que existe sobre un problema específico; principalmente, porque son menos susceptibles a sesgos e imprecisiones, gracias al protocolo con metodología explícita y rigurosa[8]. La **tabla 20.1** presenta las principales diferencias entre las revisiones sistemáticas y las revisiones narrativas.



Las RSL presentan grandes ventajas sobre las revisiones narrativas en el momento de evaluar la evidencia que existe sobre un problema específico; principalmente, porque son menos susceptibles a sesgos e imprecisiones, al seguir el evaluador un protocolo preestablecido con una metodología explícita y rigurosa.

Una RSL puede o no completarse con un metaanálisis, el cual es una técnica de análisis estadístico que permite obtener un estimado único (la estimación agrupada del efecto) a partir de los resultados de los estudios independientes (primarios). Si bien toda RSL debe regirse por los principios metodológicos que se describen en este capítulo, en algunas ocasiones (por diferencias clínicas o metodológicas de los estudios que la conforman, o por el bajo número de estudios encontrados) puede ser inapropiado, o, incluso, desorientador, hacer un metaanálisis con los datos de los estudios primarios[4]. En resumen, una RSL no siempre incluye un metaanálisis; sin embargo, todo metaanálisis debería ser parte de una RSL.

TIPOS DE REVISIONES SISTEMÁTICAS DE LITERATURA (RSL)

Este capítulo se enfoca, principalmente, en RSL que evalúan intervenciones clínicas, si bien pueden emplearse para la revisión de cualquier tipo de literatura biomédica, como estudios experimentales[7], observacionales[9] o de pruebas diagnósticas[10]. Existen RSL sobre la efectividad de intervenciones, sobre la precisión de una prueba diagnóstica o sobre el efecto de factores pronósticos y de riesgo. Todas esas revisiones, aunque comparten, en general, la misma metodología para la búsqueda, la selección, la apreciación y la síntesis de la evidencia,

Tabla 20.1 Diferencias entre las revisiones sistemáticas y las narrativas

Aspecto	Revisión narrativa	RSL
Pregunta de investigación.	Usualmente, no hay, o es amplia y poco enfocada.	**Precisa; suele centrarse en un problema** clínico o de investigación.
Sección de materiales y métodos.	Usualmente, no la hay.	Describe en detalle los procedimientos, lo que permite replicarlos.
Búsqueda de la literatura.	No se describe, o no sigue un patrón establecido.	**Pretende ser exhaustiva; emplea criterios** explícitos.
Criterios de selección de los estudios primarios.	**No son claros.**	**Explícitos y uniformes.**
Evaluación de la calidad de la información.	**No es explícita ni rigurosa.**	**Explícita; emplea métodos establecidos.**
Síntesis de la información.	Narrativa.	Narrativa, y, a menudo, cuantitativa.

presentan variaciones en la aplicación de esta metodología, propias del tipo de estudios que las conforman[11].

Los procesos para seguir una metodología sistemática en cada tipo de revisión de literatura biomédica se encuentran en diferentes niveles de desarrollo; los métodos para conducir RSL que responden preguntas sobre la efectividad de las intervenciones están mejor desarrollados que para otro tipo de preguntas, mientras los métodos de síntesis de evidencia en las RSL sobre riesgo y pronóstico apenas si empiezan a desarrollarse. Es pertinente, entonces, presentar de manera breve algunas de las consideraciones más relevantes y controversiales respecto a los retos metodológicos que enfrentan los revisores (autores de RSL) para evaluar, resumir y sintetizar la información en los diferentes tipos de RSL.

Los métodos para conducir RSL que responden preguntas sobre la efectividad de las intervenciones están mejor desarrollados que para otro tipo de preguntas.

Como se verá más adelante, uno de los factores que limitan la calidad de las conclusiones de una RSL es la calidad de los estudios que se pretende resumir. Como el experimento clínico aleatorizado (ECA) ofrece un mayor control de los sesgos de confusión, hay mayor confianza cuando la RSL se limita a ese tipo de estudios que cuando se emplea para otros diseños[7]. Sin embargo, no se puede desconocer que los estudios observacionales (por ejemplo, de cohortes o de casos y controles) son una fuente valiosa de información cuando no se pueden o no se deben realizar experimentos. Son varias las áreas de la salud en las cuales la evidencia sobre la efectividad de las intervenciones se basa, en su mayoría, en estudios no experimentales; entre ellas se cuentan la cirugía, la salud pública y la organización de los sistemas de salud, entre otras. Por ello, se ha incrementado el interés en definir métodos explícitos para la incorporación de este tipo de estudios no experimentales en las RSL de intervenciones[11]. Aunque dicha propuesta aún genera controversias, los autores de RSL que planeen responder preguntas donde se anticipan estas situaciones podrán revisar dichas propuestas metodológicas[12].

Por otra parte, las decisiones sobre el uso de intervenciones médicas deberían fundamentarse en la evaluación del balance entre los beneficios y los riesgos o los daños de estas[13]; sin embargo, los investigadores que hacen RSL sobre riesgo o daño enfrentan el desafío de obtener información pertinente y válida[14]. Si bien esas RSL se basan, principalmente, en los ECA publicados, es sabido que la identificación de las reacciones adversas importantes de un tratamiento y la cuantificación de su riesgo a menudo requieren una gama más amplia de fuentes de datos; fundamentalmente, los estudios de fase IV y los estudios posmercadeo, entre los cuales, a su vez, se incluyen estudios experimentales no publicados y estudios observacionales y de farmacovigilancia, así como información proveniente de los estudios presentados por la industria farmacéutica como parte de la solicitud de patentes y licencias para comercialización[12].

La síntesis de los datos requiere una cuidadosa consideración de la validez interna, de la

aplicabilidad y de las fuentes de heterogeneidad; debe recordarse que los criterios para juzgar la calidad metodológica y la validez de los estudios que informan daño (diferentes de los ECA o de los estudios observacionales de cohortes y de casos y controles) se hallan todavía en etapa de desarrollo. Existen cada vez más publicaciones al respecto, que presentan ejemplos de enfoques a los problemas metodológicos asociados a la realización de RSL sobre daño[15-17], pero aún no hay claridad sobre cuál es la aproximación de evaluación y síntesis más apropiada[12].

Por último, la evaluación de la precisión de pruebas diagnósticas ha suscitado especial interés entre los investigadores; actualmente se dispone de desarrollos metodológicos importantes que permiten no solo identificar de manera eficiente estudios sobre pruebas diagnósticas, sino aplicar criterios estandarizados para evaluar su validez, así como el uso de técnicas estadísticas de análisis de los datos que permiten obtener estimativos agrupados de la sensibilidad y la especificidad de una prueba diagnóstica.

La presentación detallada de los procesos específicos que se siguen en cada paso de una revisión sistemática aplicados a la RSL sobre la precisión de pruebas diagnósticas sobrepasa el carácter introductorio de este capítulo. Los investigadores que quieran profundizar sobre los aspectos específicos de la conducción de este tipo de RSL pueden consultar el *Manual de revisiones sistemáticas de pruebas diagnósticas*, de la Colaboración Cochrane[18], el cual es, hasta el momento, la principal fuente de información sobre dichos desarrollos metodológicos y su aplicación.

APORTES Y LIMITACIONES DE LAS RSL

Como se ha presentado hasta aquí, las revisiones sistemáticas de la literatura presentan un resumen del estado del conocimiento en un tema específico, dentro de un marco de referencia explícito y siguiendo métodos reproducibles. Esto puede ser de gran utilidad para quienes deben utilizar dicha información para la toma de decisiones[19] **(tablas 20.2 y 20.3)**.

Al combinar resultados de diversos estudios, las revisiones sistemáticas ofrecen mejor precisión a la hora de estimar el tamaño de un efecto y más posibilidades de generalizar los hallazgos que los estudios individuales; además, establecen las variaciones (de magnitud, o, incluso, de dirección) de dicho efecto en diversas poblaciones, estiman la frecuencia con la cual se presenta un desenlace en diferentes estudios y poblaciones, y describen el peso que un estudio determinado tiene en el contexto global de la evidencia disponible sobre un tema. Este es el principal propósito de la Colaboración Cochrane: producir revisiones sistemáticas de todas las intervenciones disponibles para el cuidado de la salud[20].

A diferencia de lo que ocurre con las revisiones tradicionales, la descripción explícita de los métodos empleados en la RSL les permite a los lectores establecer las potenciales limitaciones u omisiones de la revisión. En tal sentido, las RSL pueden contribuir a una más fácil y mejor actualización del conocimiento profesional.

Por otra parte, las RSL pueden ser inmejorables fuentes de nuevas preguntas de investi-

Tabla 20.2 Fortalezas de las revisiones sistemáticas de la literatura

• Resumen del estado del arte en un momento determinado.
• Cuando el metaanálisis es pertinente, facilitan determinar estimativos precisos de la magnitud de efectos y de variaciones en estos efectos en diferentes poblaciones.
• Permiten explorar causas de variación (heterogeneidad) entre diversos estudios, en los efectos de intervenciones en salud y en los riesgos basales.
• Facilitan estudiar el impacto y las contribuciones de un estudio determinado en el contexto global del tema.
• Son menos susceptibles a los sesgos que las revisiones narrativas.
• Permiten una aproximación más objetiva a la información.
• Ayudan a identificar nuevas preguntas de investigación.
• Facilitan el trabajo de los investigadores y de las agencias financiadoras al resaltar las brechas en el conocimiento y los estándares utilizados en la investigación existente.
• Son susceptibles de convertirse en publicaciones vivas, que se actualizan en la medida en la que se genera nuevo conocimiento.

Tabla 20.3 Limitaciones de las revisiones sistemáticas de la literatura

• Son estudios observacionales, con susceptibilidad a sesgos de selección y de medición de la calidad o del efecto de los estudios.
• Pueden presentar sesgos de publicación.
• Los hallazgos pueden tener su validez comprometida cuando los estudios son de pobre calidad.

gación, ya que hacen evidentes los vacíos de conocimiento en un área. Al describir de manera más precisa el tamaño del efecto de una intervención, ofrecen a los investigadores mejores elementos para planear aspectos específicos de sus nuevos estudios, tales como las características de la intervención, los desenlaces de mayor interés o el tamaño de la muestra. Las revisiones sistemáticas también permiten evaluar la consistencia de los estudios y explorar las principales fuentes de variabilidad en estudios que presentan resultados aparentemente contradictorios, al igual que conocer las técnicas de medición utilizadas previamente por los investigadores.

Uno de los mayores beneficios de las revisiones sistemáticas es identificar las preguntas clínicas que han sido suficientemente estudiadas y no requieren más investigación. Cada vez es más común que los investigadores que envían aplicaciones para la financiación de estudios deban justificar su propuesta a la luz de una RSL capaz de establecer si el proyecto es realmente novedoso y relevante, y evitar así la duplicación de esfuerzos y la pérdida de recursos.

EJEMPLO

El estudio de Antman, ilustra varios casos en los que se siguieron conduciendo ensayos controlados aleatorios en circunstancias donde la evidencia acumulada, si hubiera sido resumida mediante una revisión sistemática, habría bastado para tomar decisiones clínicas y suspender la investigación en el campo[3]. La mitad de los pacientes que participaron en los estudios posteriores fueron sometidos, innecesariamente, a intervenciones inferiores a las intervenciones con las que se las había comparado.

Las RSL son particularmente susceptibles de convertirse en "publicaciones vivas", que se actualizan en la medida en que se va generando nuevo conocimiento. No es inusual que las conclusiones de las revisiones varíen en respuesta a las nuevas evidencias, o que la estructura de las revisiones se vaya adaptando en la medida en que se aprende más sobre la metodología de revisiones sistemáticas.

No obstante lo anterior, como cualquier otro diseño de investigación, las revisiones sistemáticas son susceptibles de sesgos[21]. De hecho, conviene considerar que las RSL son estudios observacionales (similares a una encuesta o a un estudio de corte transversal) que tienen por sujeto de la investigación a los estudios que han evaluado la pregunta de interés. El hecho de que los autores señalen qué estrategias utilizaron para minimizar los sesgos no es garantía de que no se hayan podido cometer errores sistemáticos al seleccionar o al evaluar la literatura relevante.

La inclusión de estudios con sesgos o errores aleatorios en una revisión sistemática puede conducir a que la revisión llegue a conclusiones erradas. Adicionalmente, una de las dificultades que se presentan con las revisiones sistemáticas es que los estudios primarios incluidos pueden no haber evaluado de forma homogénea los desenlaces. Esto se conoce como ***heterogeneidad***, y es una de las razones por las cuales no siempre a una RSL le sigue un metaanálisis.



Como cualquier otro diseño de investigación, las revisiones sistemáticas son susceptibles de sesgos.

EJEMPLO

Un estudio que evalúa los efectos de un tratamiento para la anemia en las mediciones de hemoglobina puede presentar los resultados como *promedios* de hemoglobina, mientras otro los presenta en *categorías*. Más aún, los estudios que reporten categorías pueden haber utilizado puntos de corte diferentes.

Este problema es particularmente notorio cuando se evalúan efectos adversos. A menudo, estos se agrupan de forma poco clara, y los criterios para establecer su severidad son ambiguos.

EJEMPLO

Un estudio puede agrupar la cefalea inducida por analgésicos como "efecto adverso menor", en tanto que otro la clasifica como un *efecto adverso neurológico*. Al establecer el resumen de los efectos adversos, es imposible agruparlos de forma que tengan un significado clínico claro.

Finalmente, hay diversas modalidades de un error sistemático conocido como ***sesgo de publicación***: los estudios con resultados "positivos" (los que hallan diferencias estadísticamente significativas entre las intervenciones) tienden a ser aceptados más fácilmente para publicación que los estudios "negativos". Esta fuente de error es una de las más importantes para todos los estudios de carácter integrativo.

Los estudios con resultados "positivos" (los que hallan diferencias estadísticamente significativas entre las intervenciones) tienden a ser aceptados más fácilmente para publicación.

Debe mantenerse siempre un sano escepticismo para con este diseño, al igual que con cualquier otro. Hay guías para leer críticamente las revisiones de la literatura, las cuales permiten identificar sus principales fuentes de error[22-24]. Las RSL son *una fuente más* de información, y no reemplazan de ninguna manera el sentido común del clínico[25].

La **tabla 20.4** presenta los principales pasos que se deben seguir al adelantar una RSL[5]. Los párrafos que siguen describen los principios metodológicos para planear y conducir cada una de dichas fases. Al igual que con otros diseños de investigación, sean primarios o integrativos, es necesario consignar en un protocolo los procedimientos que el revisor piensa seguir para conducir la revisión.

PASOS PARA HACER UNA RSL

Definición de la pregunta de investigación

Los elementos para definir apropiadamente una pregunta de investigación se describen en el **capítulo 2** "La pregunta de investigación". Al igual que para los restantes diseños, se debe establecer: la población o el tipo de sujetos; la exposición o la intervención de interés a la cual se somete a esta población; la exposición o la intervención de control contra la cual se compara la de interés, y los desenlaces o las consecuencias que se espera modificar con la exposición o la intervención[26,27].

Las RSL, al igual que para los restantes diseños, deben establecer: la población o el tipo de sujetos; la exposición o la intervención de interés a la cual se somete a la población; la exposición o la intervención de control contra la cual se compara la de interés, y los desenlaces o las consecuencias que se espera modificar con la exposición o la intervención.

Una pregunta imprecisa o vaga generará dificultades al definir los diseños de estudios que se quiera incluir en la revisión, las estrategias de búsqueda de los estudios y la forma como se van a sintetizar los hallazgos. Por otra parte, el investigador puede hallarse interesado en evaluar diferentes intervenciones en una condición específica o los resultados de una intervención para una condición específica en diferentes poblaciones. Al igual que en una búsqueda de la literatura hecha con otros propósitos, la pregunta debe sugerir el tipo de estudios pertinentes para contestarla. Generalmente, deben buscarse de manera primordial experimentos, otras revisiones sistemáticas y metaanálisis.

La principal indicación para una RSL (e, idealmente, para un metaanálisis) es la existencia de información conflictiva sobre una intervención. Típicamente, algunos experimentos clínicos informan resultados positivos, en tanto que otros son negativos, o tienen diferencias significativas en cuanto a la magnitud del efecto.

Otra indicación habitual es la presencia de varios ECA sobre un mismo tema, sin una síntesis que facilite una visión global, aunque esta puede ser de gran utilidad en la práctica clínica o al planear investigaciones posteriores.

También es una indicación la actualización de una RSL cuando se sepa que han surgido estudios nuevos capaces de alterar las conclusiones.

La Colaboración Cochrane ofrece, a través de sus grupos de revisión (http://www.cochra-

Tabla 20.4 Pasos de una RSL

1. Plantear la pregunta de la revisión.
2. Definir los criterios de selección (inclusión y exclusión) de los estudios: • Participantes. • Intervenciones y comparaciones. • Desenlaces. • Diseño y calidad metodológica.
3. Búsqueda de los estudios: • Estudios publicados. • Estudios no publicados.
4. Selección de los estudios.
5. Evaluación de la calidad de los estudios.
6. Extraer la información necesaria de los estudios.
7. Análisis de la información: • Evaluación de la heterogeneidad. • Síntesis de resultados: medidas de resumen, modelos de efectos fijos o aleatorios. • Análisis del sesgo de publicación. • Análisis de sensibilidad.
8. Interpretación de los resultados.

ne.org/), una lista temas para los cuales sus editores consideran necesario adelantar una revisión sistemática. Con el fin de investigar sobre esos temas, basta con escribirle a la persona coordinadora del grupo expresándole el interés en conocer los temas que requieren revisiones sistemáticas.

Criterios de inclusión y de exclusión

En una revisión sistemática, el investigador trata de resumir la evidencia existente sobre un tema; por ello, el sujeto primario de tal tipo de investigación son los estudios que han evaluado una intervención, una exposición, una prueba diagnóstica o el pronóstico de una condición. Al igual que con otros tipos de investigación, el autor de una revisión sistemática debe especificar los criterios de elegibilidad de los estudios que ingresará a su investigación.

Al establecer los criterios de inclusión se debe buscar el balance apropiado entre los extremos de ofrecer criterios muy amplios o muy estrictos[28,29].

Los criterios de inclusión amplios hacen que los resultados de la revisión sean fácilmente generalizables, y, probablemente, implican que se pueda incluir en la revisión un mayor número de estudios. Sin embargo, ello conduce a que se incorporen estudios adelantados con poblaciones muy diversas, lo cual puede acabar comprometiendo la validez de la revisión, pues las comparaciones y las adiciones de los datos no tienen sentido clínico ni epidemiológico. Adicionalmente, cuando las preguntas son amplias y las poblaciones son heterogéneas, la recuperación y la síntesis de la información son más dispendiosas y más costosas. Si se tiene evidencia de que diferentes condiciones raciales, epidemiológicas o culturales pueden modificar los resultados, se debe tratar de ajustar los criterios de inclusión o hacer análisis separados de dichas poblaciones.

Por otra parte, utilizar criterios más estrictos restringiendo la revisión a poblaciones muy específicas favorece que los datos sean fácilmente "agregables" y comparables. Sin embargo, tales revisiones arrojan información aplicable a pocas personas; además, puede ser difícil encontrar un número adecuado de estudios con datos suficientes para aportar conclusiones realmente útiles.

El planteamiento apropiado de la pregunta de investigación descrito antes ofrece elementos que facilitan definir cuáles deben ser los criterios de inclusión y de exclusión para la revisión. Enseguida se describen los elementos que se deben tomar en cuenta para definir la pregunta de una revisión sistemática.



En una revisión sistemática, el investigador trata de resumir la evidencia existente sobre un tema; por ello, el sujeto primario de tal tipo de investigación son los estudios que han evaluado una intervención, una exposición, una prueba diagnóstica o el pronóstico de una condición.
Al igual que con otros tipos de investigación, el autor de una revisión sistemática debe especificar los criterios de elegibilidad de los estudios que ingresará a su investigación.

Población de estudio

Se debe definir el tipo de participantes admitidos en los estudios que se vayan a analizar en la revisión. Eso implica establecer las características específicas que se quieren identificar en los análisis (edad, raza, sexo, nivel de escolaridad y otras condiciones de las personas).

EJEMPLO

En una revisión sobre la utilidad de surfactante pulmonar en neonatos prematuros, el investigador puede estar interesado en evaluar el impacto de la intervención en neonatos prematuros de diferente peso, así como en neonatos con enfermedades asociadas o que tengan un determinado antecedente gestacional (diabetes gestacional).

Condición que se va a estudiar

Los criterios diagnósticos que se van a considerar para definir la condición o la enfermedad que se va a estudiar deben presentarse de forma explícita. Si se escogen numerosos criterios diagnósticos, posiblemente, al final no se puedan identificar estudios que cumplan con todos ellos. Los criterios muy laxos o escasos pueden acabar comprometiendo la validez del estudio, pues se incluirá a grupos de pacientes de condiciones diferentes. Desafortunadamente, no hay una fórmula mágica para definir "el justo medio". A menudo, es útil emplear definiciones pragmáticas que faciliten el consenso.

EJEMPLO

Si se quiere identificar estudios sobre la anemia en mujeres embarazadas, se puede adoptar la definición de anemia de la Organización Mundial de la Salud (OMS) (hemoglobina < 11 g/dl). Algunas personas podrían considerar muy simple tal definición, y sugerir mediciones de ferritina, aspirados de médula ósea, etc. Probablemente, al hacerlo descubran que la disparidad entre los métodos diagnósticos utilizados en diferentes estudios es tal que no tendrán material para incorporar en la revisión.

Intervención que se va a estudiar

EJEMPLO

En Colombia no es usual la administración de hierro parenteral a las mujeres embarazadas con anemia. Sin embargo, en África esta es una de las vías más comunes de administración. En la revisión, el investigador debe definir qué dosis, qué presentaciones y qué formas de aplicación de una intervención va a considerar como criterios de selección de estudios.

Se debe especificar qué intervenciones, exposiciones, pruebas diagnósticas o factores pronósticos van a ser evaluados. En el caso de las intervenciones, estas se suministran de diferentes formas y en diversos ambientes, y puede convenir que se incluya en la revisión el mayor espectro posible de opciones.

Debe tenerse presente que en las evaluaciones de intervenciones psicosociales como las terapias de apoyo puede ser difícil estandarizar la forma de administración, o que una estandarización puede resultar muy lejana a la realidad. Sin embargo, si se quiere que los resultados sean interpretables, debe buscarse una alternativa para tipificar las intervenciones.

EJEMPLO

Si se quiere comparar el impacto del uso profiláctico de surfactante en la prevención de la mortalidad neonatal con el impacto de su empleo como rescate en neonatos prematuros con dificultad respiratoria, puede ser razonable evaluar el desenlace solo en neonatos que tengan un puntaje inicial de Apgar < 7, un punto de corte usual para el clínico que quiere identificar asfixia neonatal. Hacer comparaciones discriminando por cada valor de Apgar no tiene sentido práctico ni clínico.

Desenlaces de interés

Hasta hace poco tiempo, se pensaba que incorporar todos los resultados que puedan ser significativos para los pacientes y para quienes toman las decisiones en salud enriquecía la revisión. La experiencia ha mostrado que en la práctica clínica, así como en las políticas públicas, la toma de decisiones se basa en un número reducido de desenlaces; generalmente, dos o tres de los considerados importantes por los pacientes o por los clínicos. Por esto, se recomienda un proceso explícito para seleccionar y definir un número reducido de desenlaces sobre los cuales se harán la evaluación y la síntesis de la evidencia. Como se presentará más adelante, la Colaboración Cochrane propone limitar la evaluación y la síntesis de la evidencia a un máximo de seis desenlaces, que hayan sido definidos como críticos o prioritarios tanto por los pacientes como por los clínicos y los tomadores de decisiones en salud[30].

Idealmente, la efectividad de una intervención o la precisión de una prueba diagnóstica deben evaluarse por su efecto sobre desenlaces relevantes como curación completa, superviven-

cia libre de progresión, cambios en el estado de gravedad, frecuencia y recurrencia de síntomas, tiempo para el evento (por ejemplo, siguiente estado de enfermedad, recurrencia o recaída), efectos secundarios y desenlaces adversos[31].

EJEMPLO

En una RSL sobre tratamientos parenterales para la anemia ferropénica deberán incluirse como desenlaces las reacciones alérgicas que el tratamiento pueda provocar. Estos resultados se pueden agrupar con criterios explícitos en diferentes categorías, tales como reacciones graves que comprometen la vida, reacciones graves que no comprometen la vida, reacciones menores, etc.

Diseño y calidad metodológica de los estudios que se incluirán

Los diferentes diseños de investigación sirven para contestar diferentes tipos de preguntas. Las revisiones sistemáticas, como el estudio agregativo, han tenido su mayor desarrollo en la evaluación de intervenciones en salud, pues, según se considera, los estudios primarios que las alimentan (los experimentos clínicos) alcanzan un mayor control de las variables de confusión a través de la aleatorización. Este método es la única forma de controlar los factores de confusión desconocidos o que no se han cuantificado, al lograr que su distribución sea similar en los grupos de estudio; o, al menos, si hay diferencias, que se deban al azar, y no a un sesgo. Por ello, a menudo los autores de una revisión sistemática de intervenciones deciden incluir solo experimentos clínicos. Como ya se mencionó, en casos excepcionales donde se evalúan condiciones que tienen un curso absolutamente predecible, o en los que los efectos de una intervención son muy notables, podría ser adecuado incluir estudios no experimentales.

BÚSQUEDA DE LOS ESTUDIOS

Para asegurar la validez de una RSL, se debe garantizar que se incluye en ella toda la evidencia disponible. Hacer una revisión a partir de tan solo una parte de los estudios puede llevar a que se omitan, de manera sistemática, los que muestran más a menudo resultados a favor (o en contra) de la intervención. Si eso ocurre, las conclusiones de la revisión no serán válidas. Lo anterior implica que se deben hacer todos los esfuerzos posibles por llevar a cabo una búsqueda exhaustiva y actualizada de los estudios relevantes para la pregunta de la revisión. Eso conlleva buscar tanto lo que se ha publicado como lo que se ha hecho, pero no se ha publicado[32-34].



Para asegurar la validez de una RSL, se debe garantizar que se incluye en ella toda la evidencia disponible. Hacer una revisión a partir de tan solo una parte de los estudios puede llevar a que se omitan, de manera sistemática, los que muestran más a menudo resultados a favor (o en contra) de la intervención.

Búsqueda de los estudios publicados

La búsqueda de los estudios publicados es, sin lugar a dudas, más simple que la de los no publicados. Esto, sin embargo, no significa que sea fácil ubicar todos los estudios publicados y relevantes para la revisión. Hay dos factores que lo impiden: 1) *ninguna* base de datos de literatura biomédica contiene todas las revistas biomédicas (ni, por ende, todos los estudios publicados), y 2) los estudios registrados en las bases de datos pueden estar mal clasificados.

En este momento hay centenares de bases de datos de literatura biomédica. Las más conocidas incluyen MEDLINE, EMBASE y la Biblioteca Cochrane. MEDLINE, producida por la Biblioteca Nacional de Medicina de los Estados Unidos, contiene el registro electrónico de los artículos publicados a partir de 1966 en cerca de 4000 revistas publicadas en 40 idiomas. Su consulta, a través del buscador PubMed, está disponible en Internet sin costo para los usuarios. EMBASE, por su parte, es un registro producido desde 1974 por un proveedor comercial (Elsevier) que contiene referencias de 4000 revistas. No cuenta con acceso libre a través de Internet. Se estima que estas dos bases de datos tienen una sobreposición del 70 %. Por último, la Biblioteca Cochrane incluye el Registro Cochrane de Experimentos Clínicos (Cochrane Central Register of Controlled Trials —CENTRAL—), la base de datos más completa sobre este tipo de estudios, con cerca de 400 000 experimentos.

Adicionalmente existen bases de datos organizadas por áreas temáticas (cáncer, toxico-

logía, SIDA, salud reproductiva) o por regiones (LILACS, para la literatura latinoamericana). El lector interesado puede encontrar más información sobre estos y otros recursos en el **capítulo 3** "Búsqueda de la literatura médica", en este mismo libro.

Como no todas las revistas forman parte de las bases de datos mencionadas, conducir la búsqueda en solo una o dos de ellas no garantiza que se obtenga la totalidad de los estudios publicados. Una búsqueda conducida en MEDLINE por un experto logra identificar cerca del 50 % (rango del 17 %-82 %) de los estudios sobre el tema que se hayan publicado[35]. Esto parecería razonable si se acepta que MEDLINE solo contiene una baja proporción (cercana al 15 %) de las revistas biomédicas que se estima que se publican en el mundo. Por otra parte, este hallazgo tiene grandes repercusiones para quien quiera adelantar una RSL: limitar la búsqueda a una sola o a pocas bases de datos, así sean las más completas, seguramente llevará a dejar fuera de la revisión parte de los estudios publicados. El revisor debe tratar de hacer su búsqueda en varias bases de datos. No existe una cifra mínima: cuantas más, mejor.

La situación mejora un poco cuando el experto limita su búsqueda solo a las revistas que están incluidas en MEDLINE: se logra identificar el 77 % (rango del 32 %-91 %) de los artículos que existen sobre el tema en la base de datos[23]. Ello demuestra la segunda gran limitación de las bases de datos de la literatura: muchos artículos están mal clasificados, lo cual significa que no es posible recuperarlos de la base de datos, aunque se encuentren en ella. De ahí que la búsqueda en más de una base de datos pueda mejorar la probabilidad de encontrar estudios.

También se recomienda que el revisor busque artículos relevantes en las referencias de los que ya ha obtenido (búsqueda secundaria). Es un buen sistema para identificar nuevos estudios, aunque puede favorecer los sesgos de referencia, al aumentar la probabilidad de incluir estudios cuyos resultados se apoyan entre sí. También se deben revisar las referencias de las revisiones publicadas (sistemáticas o narrativas).

Búsqueda de los estudios no publicados

Está demostrado que es más común la publicación de estudios con hallazgos positivos (en los cuales una intervención es mejor que otra) que la de estudios con resultados negativos[11]. Más aún, los estudios positivos tienden a publicarse más tempranamente, a publicarse más a menudo en inglés y a citarse más que los que muestran resultados negativos. Todo esto puede llevar a una situación conocida como *sesgo de publicación*: si se incluyen en una RSL solo estudios publicados, eso puede hacer que haya una mayor representación de estudios positivos que de negativos, lo cual, a su vez, puede llevar a que se sobrevalore el efecto de una intervención.

Por supuesto, las dificultades para localizar estudios no publicados son mayores que las descritas antes para encontrar los publicados. Aun si se usan estrategias complementarias, no habrá garantía de que se han localizado todos los estudios no publicados. Por ello, la RSL debe incluir en el análisis la medición de la probabilidad de sesgo de publicación. Las estrategias para ello se describen más adelante en este capítulo.

La solución al sesgo de publicación sería la creación de registros donde los investigadores inscribieran sus experimentos, incluso cuando se encuentran en planificación o en ejecución. La consulta de dichos registros puede llevar a identificar experimentos adicionales que aporten mayor objetividad a la revisión. Ya hay varios registros especializados para diferentes enfermedades, aunque la mayoría de los existentes se limitan a experimentos clínicos controlados[20].



La solución al sesgo de publicación sería crear registros donde los investigadores inscribieran sus experimentos, incluso cuando se encuentran en planificación o en ejecución.

Otra estrategia para buscar estudios no publicados es comunicarse con los investigadores que han hecho publicaciones en el tema, y preguntarles si conocen estudios no publicados que sean relevantes para la revisión. Tal estrategia es cada día más fácil de seguir, gracias a la disponibilidad de sistemas de búsqueda y de los correos electrónicos de los autores de los artículos publicados, y que se consiguen a menudo en MEDLINE, por ejemplo. La búsqueda de estudios no publicados puede también extenderse a expertos locales o internacionales en el tema, o a las compañías que han desarrollado o comercializan el producto de interés para la revisión.

Cuando se escribe el protocolo o el informe de la RSL, los revisores deben describir la estrategia de búsqueda utilizada, y hacerlo con el suficiente detalle para que otro investigador pueda reproducir la búsqueda. Para eso, es necesario reportar las bases de datos consultadas, los términos y los conectores empleados, los años incluidos y si se utilizaron filtros o límites (por idioma, por ejemplo, por años de publicación, etc.). También se deben describir las otras fuentes de información empleadas; en particular, los esfuerzos para localizar estudios no publicados.

Es común que en ocasiones se intente restringir la búsqueda a los artículos disponibles en las bibliotecas locales o a las publicaciones hechas en los idiomas que el revisor domina. Este tipo de restricciones introduce sesgos y compromete de forma importante la validez del estudio[20]. Para evitar eso, el investigador debe considerar la posibilidad de obtener la traducción de las secciones de resultados y de materiales y métodos de los artículos publicados en otros idiomas, y que parezcan pertinentes a la revisión. Afortunadamente, cada día es más fácil obtener artículos de publicaciones que solían estar fuera de alcance, gracias a los sistemas de informática en las bibliotecas de muchos centros de investigación.

Por último, es posible que al finalizar la búsqueda de la literatura el revisor haya identificado uno o varios estudios pertinentes a la revisión y que aún no han concluido. Es necesario, entonces, describir los estudios en curso que en el futuro servirán para actualizar la revisión. Se identifican, junto con la intervención que evalúan, la definición de desenlaces que utilizan y los datos de las personas que deben ser contactadas para obtener los resultados.



Es necesario, como parte del informe, describir los estudios en curso que en el futuro servirán para actualizar la revisión.

MANEJO DE LOS ESTUDIOS RECUPERADOS

Como es probable que para la revisión se encuentre un gran volumen de referencias, es importante que el revisor defina en el protocolo la estrategia que va a emplear para su manejo. Este es uno de los puntos que pueden facilitar o dificultar enormemente el trabajo de un revisor.

Hay diferentes formas de organizar las referencias. La primera es de forma manual, con algo tan sencillo como un fichero donde se escriben de forma estandarizada los datos de cada referencia. Estas fichas se agrupan inicialmente como referencias generales, o artículos que soportan los antecedentes de la revisión, o como presentaciones identificadas para revisión. Al iniciar la búsqueda, las fichas de artículos de revisión deberán agruparse en tres compartimientos separados:

- El primero, identificado como ***referencias solicitadas***, incluirá las referencias solicitadas a la biblioteca o al proveedor de artículos. Es recomendable escribir en la tarjeta la fecha de la solicitud, para llevar un control adecuado de la bibliografía solicitada y recuperada.
- El segundo compartimiento debe tener las fichas de las publicaciones que el revisor ya ha obtenido. Puede marcarse como ***referencias obtenidas***.
- El tercero incluirá las publicaciones que no se ha logrado obtener, y que se intentará ubicar mediante otros sistemas. Estas pueden identificarse como ***referencias no ubicadas***.

Una vez organizados estos casilleros y adelantado el trabajo de selección de las publicaciones (que se describe más adelante), las fichas de los artículos recuperados y que van a ser sometidos a la evaluación de calidad se organizarán en tres nuevos compartimientos: 1) ***estudios que deben incluirse en la revisión***, 2) *estudios que deben excluirse de la revisión* y 3) ***artículos que requieren datos adicionales o*** *mayor análisis antes de tomar una decisión*. Para tomar una decisión sobre estos últimos, suele ser necesario contactar a los autores. La disposición de los autores para colaborar es excelente a veces, y el esfuerzo invertido en ellos se justifica plenamente.

Otra alternativa para hacer la tarea de registro de referencias es utilizar un programa de computador. Hay para eso diferentes opciones; entre las más utilizadas por revisores están Reference Manager©, ProCite©, EndNote©, Mendeley© y RefWorks©. Algunas de ellas requieren suscripción y pago para su uso, y otras son de libre

acceso o son accesibles a través de licencias institucionales y corporativas. Los revisores deben consultar a la biblioteca de su institución sobre cuál está disponible o es la apropiada en su ámbito de trabajo o en la plataforma de su preferencia.

SELECCIÓN DE LOS ESTUDIOS

Tras la búsqueda de los estudios, el revisor, seguramente, tendrá una lista extensa de referencias, a partir de la cual deberá seleccionar las que incluirá en su revisión. Si la búsqueda fue realmente exhaustiva, lo más probable es que dicho listado contenga la mayor parte de los estudios que son relevantes (idealmente, todos), mezclados con un número mucho mayor que no lo son. El desafío en este punto es cómo separar ambos grupos de forma eficiente y sin introducir sesgos[36].

El proceso de selección implica seguir varios pasos de forma iterativa, para reducir de manera gradual el número de estudios. En una primera fase, quizás se parta de la revisión de los títulos o de los resúmenes, lo que permitirá separar los estudios en tres grandes categorías: 1) definitivamente excluido, 2) definitivamente incluido y 3) una tercera, donde se requiere más información para decidir. En un segundo paso, los revisores aplicarán de nuevo los criterios de elegibilidad empleando el texto completo de los estudios de las dos últimas categorías, con el objeto de decidir de forma definitiva cuáles estudios se incluyen en la revisión. Los revisores deberán hacer explícitas las razones para excluir cada uno de los artículos, dado que esta información debe hacer parte de la sección de resultados de la **RSL.**

El proceso de selección implica seguir de forma iterativa varios pasos, con el fin de reducir de manera gradual el número de estudios.

El proceso de selección de los estudios se debe hacer a la luz de los criterios de elegibilidad establecidos previamente. Dado que establecer si un determinado estudio los cumple o no se basa en un juicio, esta fase de la **RSL** puede originar sesgos si se decide incluir, de preferencia, estudios que muestren un determinado resultado. Se proponen dos estrategias para que la selección sea libre de sesgos:

La primera, más aceptada, es involucrar al menos a dos revisores en la decisión haciendo que cada uno de ellos decida, a la luz de los criterios de selección, pero de forma independiente (o sea, sin conocer uno las decisiones del otro), cuáles estudios deben incluirse o excluirse[18,19]. En caso de desacuerdo, se puede acudir a la discusión de las razones para este, con el fin de alcanzar un consenso o a la mediación de un tercer revisor que sirva como árbitro. Se recomienda, además, medir el acuerdo entre los revisores a lo largo del proceso empleando las estrategias apropiadas para ello (por ejemplo, midiendo la concordancia mediante el estadístico kappa, que se describe en el **capítulo 18** "Estudios sobre correlación y concordancia", de este libro).

Establecer si un determinado estudio cumple o no los criterios de selección se basa en un juicio, lo que puede originar sesgos; por ello, es pertinente involucrar en la decisión al menos a dos revisores haciendo que cada uno de ellos decida a la luz de los criterios de selección, pero de forma independiente.

Menos a menudo, los revisores deciden la inclusión de los estudios con base en copias de estos de las que se ha eliminado información que permita conocer sus resultados, el autor, la revista en la cual se publicó o la institución donde se condujo la investigación. Esto, que a primera vista parecería una estrategia útil para reducir la probabilidad de sesgos en la selección, no ha confirmado su capacidad para modificar los resultados del proceso. Adicionalmente, significa mayores dificultades en la preparación del material, o puede ser que esta sea, simplemente, imposible: si los revisores conocen a fondo algunos de los estudios, los reconocerán fácilmente, pese a los intentos por ocultar parte de la información.

En unos pocos párrafos o con la ayuda de una tabla, el revisor puede resumir las características generales de los estudios incluidos en su revisión. Deben describirse características generales, como el sitio donde fueron realizados, las características de los participantes, el diseño del estudio, la estrategia de aleatorización, las intervenciones evaluadas, la definición de los desenlaces, el tiempo de seguimiento, las pérdi-

das durante el seguimiento y el enmascaramiento del estudio. De la misma forma, en otra tabla, se describe el motivo por el cual se excluyó cada uno de los estudios que no ingresaron a la revisión, y se especifica el criterio de exclusión que se presentó.

EVALUACIÓN DE LA CALIDAD DE LOS ESTUDIOS INCLUIDOS

Luego de seleccionar los estudios que formarán parte de la revisión, es necesario evaluar su calidad metodológica y extraer de ellos los datos necesarios. Estos dos pasos suelen hacerse de manera simultánea. Primero se describen las estrategias para la evaluación de la calidad.

Como se ha dicho antes, una RSL es un estudio integrativo que tiene como objeto de estudio investigaciones previas. Es obvio que las conclusiones de la revisión no pueden ser de mejor calidad que las de los estudios que se revisan. En otras palabras, la calidad de los estudios primarios es uno de los principales determinantes de la validez de las conclusiones de la revisión. Cabe aquí una analogía: un procesador de basura solo produce al final basura procesada. Es por ello de capital importancia que los revisores hagan una rigurosa evaluación de la calidad de los estudios incluidos en su revisión[37]. Por otra parte, las variaciones en la calidad metodológica de los estudios puede ser una de las posibles explicaciones para la falta de consistencia (heterogeneidad) en los hallazgos de investigaciones que, aparentemente, evalúan la misma intervención.



Una RSL es un estudio integrativo que tiene como objeto de estudio investigaciones previas. La calidad de los estudios primarios es uno de los principales determinantes de la validez de las conclusiones de la revisión.

Si bien es evidente que se necesita evaluar la calidad de los estudios incluidos en la revisión, hay menos consenso respecto a los métodos más apropiados para hacerlo y a la forma de incorporar los resultados de la evaluación al tratar de adelantar un metaanálisis. Así como ocurre con otros aspectos de la metodología de las RSL, se tiene mejor información respecto a la evaluación de calidad de experimentos clínicos aleatorizados, que son los estudios más adecuados para adelantar RSL sobre la efectividad de intervenciones[25].

Si bien se han desarrollado numerosas escalas para medir la calidad de los mencionados estudios, la mayor parte de los expertos recomienda no emplearlas, pues pocas de ellas están validadas, y, particularmente, porque cuando se aplican diferentes escalas para evaluar un mismo estudio los resultados obtenidos son contradictorios: de acuerdo con algunos instrumentos el experimento fue de buena calidad mientras que otros sugieren lo contrario. De esta manera, la selección de la escala puede ser crítica en el proceso, y, más importante aún, puede dar pie a sesgos durante la revisión: un revisor podría emplear varias escalas para evaluar la calidad de los estudios de su revisión y reportar los resultados obtenidos con la que arroja los "mejores" resultados.

Por las razones anteriores, la tendencia al evaluar la calidad de los ECA es emplear criterios que han demostrado empíricamente estar asociados a posibles modificaciones en los resultados[38,39]. En consecuencia, la mayor parte de las revisiones sistemáticas se limitan a evaluar la calidad de los experimentos incluidos en ella determinando si las estrategias para producir la secuencia de aleatorización, para implementarla de manera enmascarada, para evaluar de forma ciega los resultados y para controlar la pérdida de participantes fueron adecuadas o inadecuadas, o si no se las puede evaluar debido a una descripción inapropiada. La Colaboración Cochrane propone definir "el riesgo de sesgo" de los ECA incluidos en la RSL partiendo de una serie de criterios orientados a evaluar las estrategias utilizadas por los autores para minimizar el riesgo de sesgos en cada uno de los mencionados aspectos del diseño del estudio y la conducción del estudio[26].



La tendencia al evaluar la calidad de los ECA es emplear criterios que han demostrado empíricamente estar asociados a posibles modificaciones en los resultados.

Tampoco hay claridad respecto a qué se debe hacer con el resultado de la evaluación de la calidad; en concreto, sobre cómo utilizar

el producto del proceso de esta evaluación en un eventual metaanálisis de los datos de los estudios. La mayor parte de los autores emplea dicha información para adelantar análisis de sensibilidad, los cuales se describen en detalle más adelante. En pocas palabras, se comparan los resultados del metaanálisis incluyendo todos los estudios, independientemente de su calidad, y luego, excluyendo los que tienen alguna deficiencia (más comúnmente, los que no tuvieron una aleatorización enmascarada). Si los resultados de esos dos análisis son similares, el revisor puede concluir que la calidad de los estudios incluidos tiene poco efecto sobre sus conclusiones.

MÉTODOS DE EXTRACCIÓN DE LOS DATOS

La información que debe obtenerse de cada estudio está determinada por la pregunta de interés de la revisión. En general, es necesario recuperar información respecto a las características clínicas y demográficas de la población (el número de sujetos en cada intervención, sus edades, los criterios de diagnóstico de la condición, etc.), las características de las intervenciones (el tipo, la forma de administración, la duración, la dosis en el caso de medicamentos, etc.) y los desenlaces de interés[40,41]. Este proceso es más fácil si se emplean formularios para consignar los datos que han sido probados con anterioridad. Es posible que los estudios reporten la información de interés en formas que no son idénticas: mientras algunos emplean desviaciones estándar para describir la dispersión de una variable continua, otros reportan el error estándar. Se recomienda que los revisores extraigan del formulario la información tal como se reporta en el estudio, para luego hacer las transformaciones que sean necesarias.

Es conveniente que la evaluación de la calidad y la extracción de los datos sean conducidas al menos por dos o más investigadores de forma independiente, y que cada uno de ellos permanezca enmascarado a los resultados y las conclusiones del otro. Esto reduce la posibilidad de sesgos en la evaluación de los estudios de la revisión, así como la probabilidad de errores en la extracción o la trascripción de los datos de cada estudio. Es necesario especificar de antemano cómo se resolverán las discrepancias entre los observadores (por discusión y consenso, o mediante arbitraje por una tercera persona), así como medir la concordancia alcanzada; especialmente, al evaluar la calidad de los estudios.

ANÁLISIS DE LA INFORMACIÓN

Una vez se hayan extraído los datos de interés de los estudios originales, el revisor debe decidir cuáles son las estrategias más apropiadas para resumir y analizar toda esa información. Usualmente, ello significa adelantar de forma simultánea varias tareas, que, para efectos de claridad, se describen de manera separada en las siguientes secciones. Tales estrategias de análisis estadístico (y en particular, la generación de medidas de asociación acumuladas para todos los estudios incluidos) es lo que se conoce como el ***metaanálisis de los datos***[42,43].



Las estrategias de análisis estadístico para resumir toda la información en la RSL (y en particular, la generación de medidas de asociación acumuladas para todos los estudios incluidos) es lo que se conoce como el metaanálisis de los datos.

EVALUACIÓN DE LA HETEROGENEIDAD

Una RSL pretende resumir cuantitativa o cualitativamente los resultados de múltiples estudios que, empleando diseños similares, han tratado de dar respuesta a la misma pregunta de investigación. Es intuitivamente evidente que dicho resumen es más apropiado y confiable cuanto más similares sean entre sí los estudios. Sin embargo, los estudios casi siempre tienen variaciones de diverso grado en sus métodos. Por ello, debe evaluarse el grado de heterogeneidad (o de su opuesto: la homogeneidad) de los estudios primarios. El resultado de esta evaluación es uno de los factores más importantes para establecer si se puede proceder a una síntesis cuantitativa de los resultados.

Es necesario señalar que existen dos formas básicas de heterogeneidad: la ***clínica*** y la *estadística*[44]. La primera conlleva establecer

la medida en la cual las poblaciones, las exposiciones o las intervenciones y los desenlaces de los estudios son semejantes. Esta es una decisión que se basa en comparar las características mencionadas entre los estudios empleando criterios clínicos. Por otra parte, la heterogeneidad estadística se determina haciendo pruebas que miden la probabilidad de que las diferencias observadas en los resultados de diversos estudios se deban al azar[45]. Esta es una de las pocas circunstancias donde el investigador quiere encontrar un resultado que "no sea estadísticamente significativo". Si las pruebas muestran una baja probabilidad (< 0,05), se concluye que el azar no es una explicación suficiente para las diferencias de los resultados entre los estudios, o, lo que es igual, que hay heterogeneidad de los estudios en otros factores. Por el contrario, si la probabilidad mencionada es > 0,05, el revisor puede, en principio, concluir que es probable que el azar explique las diferencias en los resultados entre los diversos estudios. Esto, a su vez, puede interpretarse como sugerencia de que no hay diferencias reales entre los estudios; o, en otras palabras, de que estos son homogéneos.



Es necesario señalar que existen dos formas básicas de heterogeneidad: la clínica y la estadística.

Otra medida de la heterogeneidad es el estadístico I^2. Dicho estadístico calcula el porcentaje de la variabilidad en el efecto observado, que puede deberse a causas diferentes del azar[6]. Como regla general, se considera que si ese porcentaje es < 30 %, la variabilidad por razones diferentes del azar es poco importante.

¿Cuál es la interpretación de los resultados? En principio, si los estudios son heterogéneos, no es buena idea que se proceda a obtener un resumen cuantitativo de los datos. Hacerlo puede carecer de sentido clínico y epidemiológico, y llevar a conclusiones incorrectas. En este caso, el revisor debe proceder a explorar las razones para la heterogeneidad de los estudios, lo que implica evaluar si hay diferencias en las poblaciones, en las intervenciones, en los desenlaces, en el diseño de los estudios o en su calidad metodológica. Una forma adicional de explorar la heterogeneidad es adelantar análisis por subgrupos de pacientes o de estudios. Dichos análisis de subgrupos deben haberse previsto desde el protocolo. Si el revisor decide, pese a la heterogeneidad, agrupar los datos en un metaanálisis, debería emplear un modelo de efectos aleatorios, como se explica más adelante.

Finalmente, se debe señalar que un resultado "no estadísticamente significativo" en las pruebas de heterogeneidad no necesariamente significa que los estudios son homogéneos. Tales pruebas tienen bajo poder; en especial, cuando el número de estudios es reducido o cuando los estudios son de tamaño pequeño. Ello significa que en dichas circunstancias, las pruebas pueden fallar a la hora de encontrar la heterogeneidad, aun cuando, de hecho, exista. El análisis clínico de los estudios cobra particular importancia en esas circunstancias: si el investigador cree que hay heterogeneidad clínica, tampoco debería proceder a adelantar el metaanálisis.

SÍNTESIS DE RESULTADOS

Si después de las consideraciones anteriores se cree que es apropiado conducir un metaanálisis, el paso siguiente consiste en evaluar un efecto global a partir de la combinación de los datos de los estudios. Ello, a su vez, tiene dos pasos: primero se obtiene una medida de resumen para cada estudio, y luego se calcula un promedio ponderado para todos los estudios. El simple promedio aritmético de los resultados de los estudios es inapropiado, pues los estudios de menor tamaño se hallan sujetos a más variabilidad por el azar que los más grandes. En consecuencia, la síntesis debe adjudicar una ponderación diferente a cada estudio, lo que, en la práctica, se hace con base en el inverso de la varianza (la que, a su vez, está muy relacionada con el tamaño de la muestra de cada estudio). Esto hace que los estudios de mayor tamaño tengan más peso en el metaanálisis.

Adicionalmente, la técnica para resumir cuantitativamente los estudios varía de acuerdo con la naturaleza de los datos analizados y con los resultados de las pruebas de heterogeneidad. Hay que hacer en este punto una diferenciación entre los modelos y los métodos del metaanálisis. Existen dos modelos: uno ***de efectos fijos*** y otro ***de efectos aleatorios***. La diferencia entre ellos radica en lo que se presume respecto a los estudios de los cuales provienen los datos. El modelo de efectos fijos presume que todos los estudios dan una estimación del "verdadero"

efecto, y que la única explicación para que sus resultados no sean idénticos es el azar. El de efectos aleatorios, por su parte, asume que los resultados de los estudios individuales son una muestra aleatoria de una "población" de diferentes resultados, no necesariamente iguales entre sí, lo cual significa que, además de la variabilidad por el azar, pueden existir otras fuentes de variación entre los estudios[5,30,31,33].

La decisión sobre cuál de los modelos se debe emplear se basa en la heterogeneidad de los estudios. Si esta no existe, los resultados obtenidos con los dos modelos son prácticamente iguales; la única diferencia podría ser una menor precisión, medida a través de los intervalos de confianza, con el modelo de efectos aleatorios. Por el contrario, si hay heterogeneidad, el modelo de efectos fijos es inadecuado, pues los datos están mostrando cómo su presunción (que la única razón para la variabilidad es al azar) no se cumple. Aplicarlo en estas circunstancias producirá resultados que difieren considerablemente de los obtenidos con el de efectos aleatorios, con intervalos de confianza artificialmente estrechos y que sugieren diferencias entre las intervenciones que realmente pueden no existir.



La técnica para resumir cuantitativamente los estudios varía de acuerdo con la naturaleza de los datos analizados y con los resultados de las pruebas de heterogeneidad. Hay que hacer en este punto una diferenciación entre los modelos y los métodos del metaanálisis.
Existen dos modelos para hacer el metaanálisis de los datos: el de efectos fijos y el de efectos aleatorios.

Si hay heterogeneidad, la literatura sugiere que el revisor debe buscar explicarla y debe considerar no combinar cuantitativamente los resultados de los estudios. Si, pese a ello, decide hacer la síntesis, debe emplear el modelo de efectos aleatorios. Otra estrategia que puede darle más solidez al análisis es estudiar los datos con los dos modelos y determinar si los resultados cambian. Si estos son iguales, el revisor tiene mayor confianza en la validez de la síntesis.

Cada uno de los dos modelos tiene, a su vez, varios métodos estadísticos. La selección final del método depende del tipo de datos (binarios, continuos, tiempo para evento) y de la medida de asociación que sea apropiada (riesgo relativo —**RR**—, riesgo relativo indirecto —RRI— o diferencia de riesgos para los datos binarios, y diferencia ponderada de medias o diferencia estandarizada de medias para los datos continuos)[5,30,31,33]. Si el revisor no tiene experiencia con el empleo de los métodos estadísticos de síntesis, es una sabia decisión pedir asesoría a un bioestadístico en el momento de definir los criterios para escoger cada modelo y al adelantar el análisis de los resultados.

Además de los métodos de síntesis cuantitativa expuestos, hay otros, como el metaanálisis de variables múltiples, los acumulativos, el análisis de supervivencia, etc. Dichos métodos no se describirán en este libro.

Presentación gráfica de los resultados

A menudo se recurre a gráficos para resumir los resultados de una revisión sistemática. El más común es la gráfica "de bosque", que muestra los hallazgos de cada estudio y la medida de resumen obtenida.

EJEMPLO

La **figura 20.1** muestra los resultados de una revisión sistemática de la Colaboración Cochrane que evaluó la efectividad de los corticosteroides para el manejo de laringotraqueítis en niños[46]. Esta figura presenta el efecto que dichos medicamentos tienen, en comparación con un placebo o con ninguna intervención, en la frecuencia de admisiones al hospital de 24-48 horas después de la intervención.
Cada cuadrado representa el RRI hallado en cada estudio. Las líneas horizontales que se extienden a ambos lados del cuadrado corresponden a la precisión de este RRI, medida con los intervalos de confianza (IC) del 95 % (IC 95 %). La línea vertical señala un RR de 1,0, lo cual indica que no hay diferencias entre las intervenciones. Los resultados ubicados a la izquierda de esta línea sugieren que la intervención reduce el riesgo del evento, mientras los localizados a la derecha sugieren que lo aumenta.
El tamaño de cada cuadrado indica el peso relativo que ese estudio tiene en el metaanálisis. El rombo representa la medida de resumen obtenida en el metaanálisis: su sección central más ancha representa el valor puntual del RRI (0,27 para el conjunto de todos los esteroides), mientras que sus extremos izquierdo y derecho señalan los límites del IC 95 % (0,18 y 0,42). Finalmente, en la parte inferior izquierda se presenta el resultado de la prueba de heterogeneidad para este desenlace (p = 0,277).

La forma gráfica más común para resumir los resultados de una revisión sistemática es la gráfica "de bosque", que muestra los hallazgos de cada estudio y la medida de resumen obtenida.

Análisis de sesgo de publicación

Como ya se dijo, la inclusión preferencial de estudios publicados en una **RSL** puede llevar a conclusiones erróneas (sesgos de publicación). Como no hay estrategias que garanticen el control de dicho sesgo, el análisis de los datos suele incluir técnicas que permiten cuantificar la probabilidad de este tipo de error.

Existen varias técnicas que se pueden emplear para tal propósito[47]. **La más conocida es** la ***gráfica de embudo* (*funnel plot*, en inglés), que** es un diagrama de puntos donde se ubica cada estudio en relación con su medida del tamaño del efecto (en el eje horizontal) y su tamaño de muestra u otra medida relacionada (en el eje vertical).

El nombre de gráfica de embudo se explica porque la precisión en la medida del efecto aumenta a medida que aumenta el tamaño de la muestra. En otras palabras, el efecto estimado por estudios pequeños es menos preciso (***varía más ampliamente*) en la base de la gráfica.** Si la gráfica es asimétrica, se presume que se "han perdido" estudios pequeños que muestran un resultado negativo, lo cual significa que, posiblemente, hay sesgo de publicación. En estas circunstancias, el revisor y los usuarios de la información tienen menos confianza en las conclusiones de la revisión.

EJEMPLO

La **figura 20.2** representa una gráfica de embudo. Los puntos representan el tamaño de la muestra y el RR de cada estudio que ha evaluado una asociación. Los puntos corresponden a los estudios publicados, mientras los círculos indican los no publicados. Una metaanálisis que incluya todos los estudios (puntos negros y blancos) produciría un RR total < 1 (línea vertical punteada), mientras otro que solo incluya los estudios publicados (puntos negros) produciría un RR aún menor (línea vertical de líneas discontinuas), y eso llevaría a sobreestimar el tamaño del efecto. La exclusión de los puntos blancos, por su parte, crearía asimetría en la gráfica.

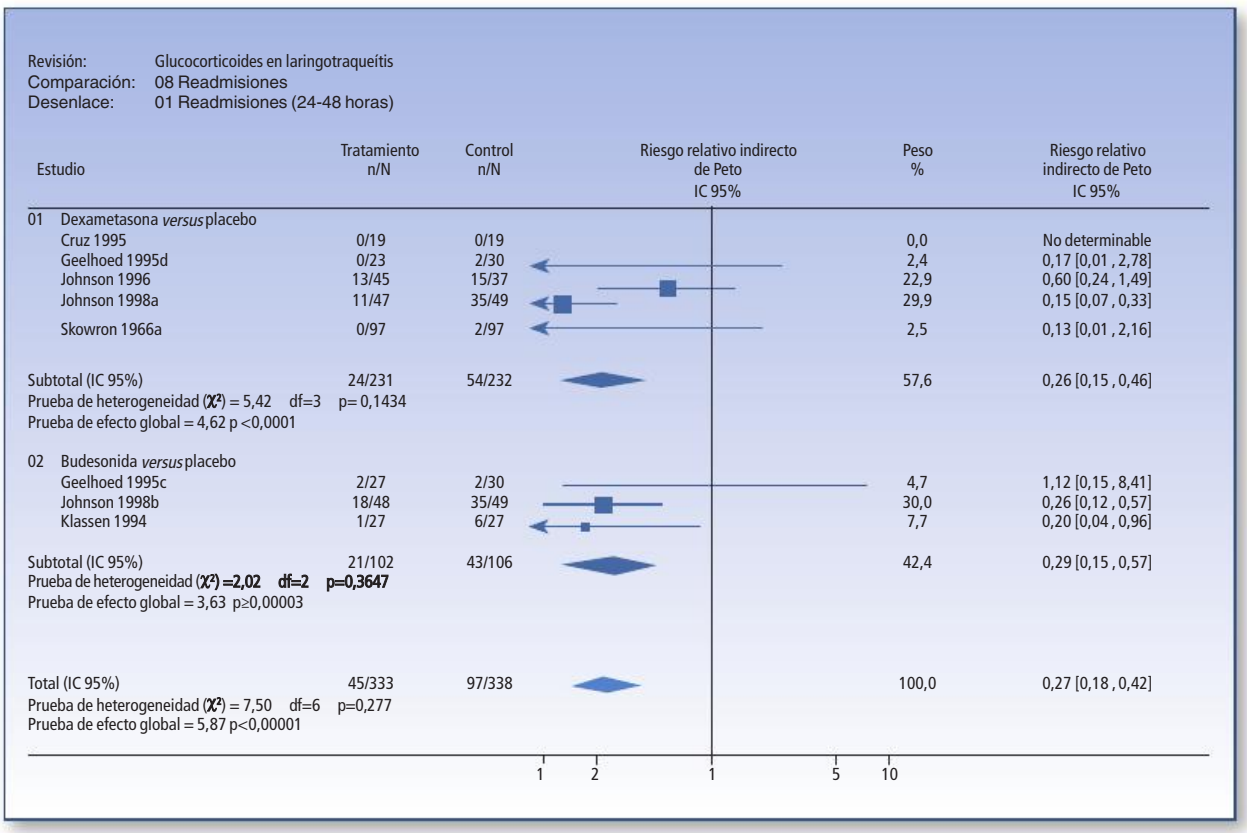


FIGURA 20.1 Gráfica de bosque de una revisión sistemática.

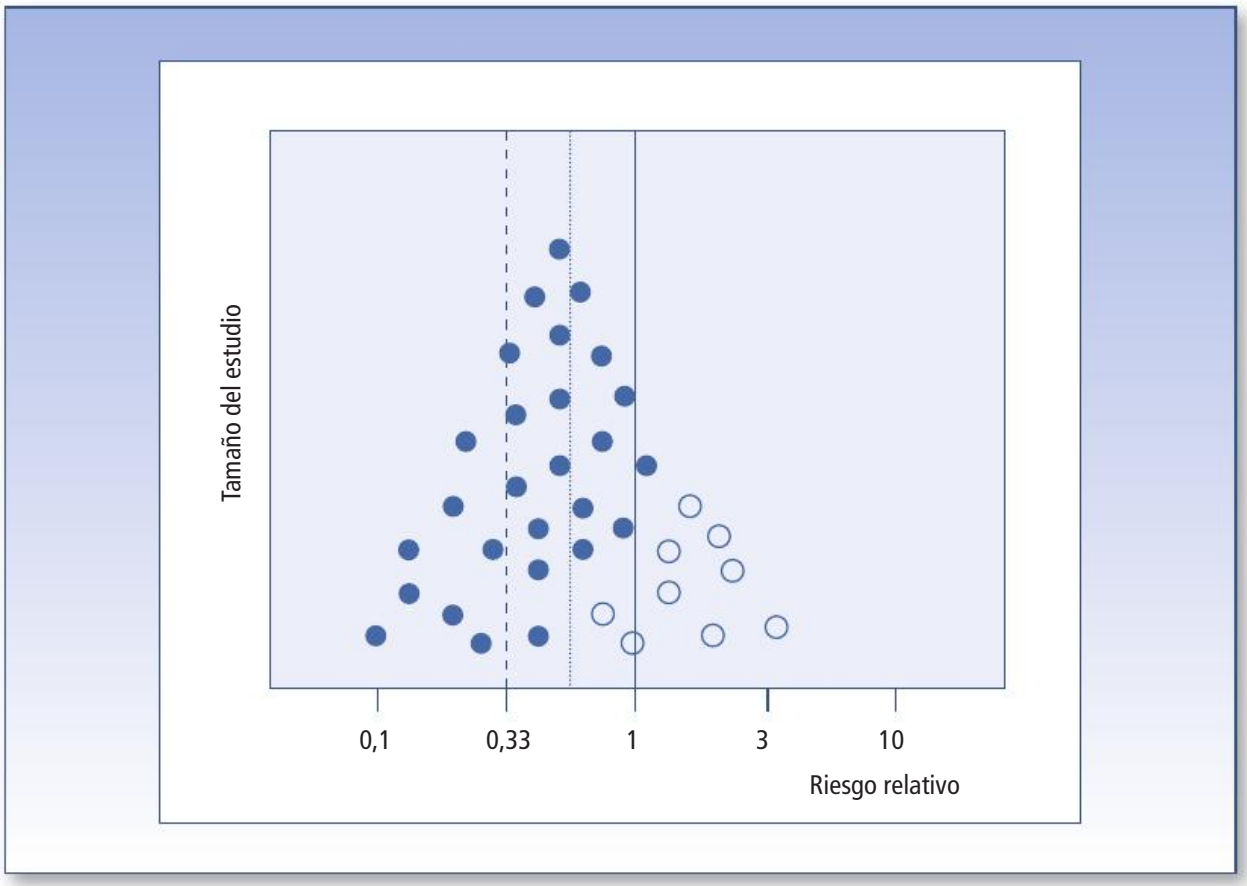


Figura 20.2 Gráfica de embudo.

Análisis de sensibilidad y de subgrupos

El análisis de sensibilidad busca explorar qué tan robustos son los resultados de la RSL ante modificaciones en algunas de sus presunciones. En principio, la técnica se basa en adelantar dos análisis diferentes variando de alguna manera los datos del estudio. Si los resultados no muestran cambios significativos, el revisor y el lector pueden tener más confianza en los resultados. Por el contrario, si los resultados varían de manera importante, es necesario proceder con cautela cuando se interpreten los resultados. Ya se han mencionado algunas circunstancias en las que se puede emplear dicha estrategia. El análisis incluyendo y excluyendo los estudios de más pobre calidad, así como la repetición de los análisis empleando los modelos de efectos fijos y los aleatorios son ejemplos de esta técnica.

La misma estrategia puede emplearse para analizar el impacto que tienen las modificaciones en las intervenciones o en las características de los sujetos del estudio. Ello corresponde más a un análisis de subgrupos, a través del cual se pretende explorar si dichas variaciones modifican la magnitud, o, incluso, la dirección del efecto.

EJEMPLO

Una RSL que evalúe el efecto de los corticosteroides en el manejo de niños con bronquiolitis puede recurrir a un análisis de sensibilidad incluyendo inicialmente todos los estudios, y luego, excluyendo los que han evaluado a niños con más de un episodio de sibilancias, y quienes, más probablemente, tienen asma en vez de bronquiolitis.

EJEMPLO

La **figura 20.1** presenta por separado los efectos de la dexametasona y de la budesonida en el riesgo de admisión hospitalaria en niños con laringotraqueítis. Se podría emplear la misma técnica para establecer si el impacto de la intervención es diferente en niños vistos ambulatoriamente, en relación con lo observado en quienes requieren hospitalización.

Es importante señalar que el número de análisis de subgrupos que se adelanten en la revisión sistemática debe ser reducido, debe contar con bases clínicas y, lo más importante, debe planearse de antemano y quedar consignado en el protocolo de la revisión. De lo contrario, sus hallazgos pueden ser desorientadores, pues las diferencias halladas pueden obedecer a la realización de múltiples comparaciones estadísticas.

EVALUACIÓN DE LA CALIDAD DEL CUERPO DE EVIDENCIA PARA CADA DESENLACE

Como la evidencia global, o cuerpo de evidencia final, obtenida en la RSL puede diferir para cada desenlace evaluado en cuanto al número de estudios que la componen, su riesgo de sesgo, la homogeneidad y la consistencia de sus resultados, el tamaño del efecto y la precisión de las estimaciones agrupadas de este efecto, el grupo "Applicability and Recomendations Methods Group", de la Colaboración Cochrane, recomienda evaluar, como parte de la RSL, la calidad de la evidencia global, o cuerpo de evidencia, para cada desenlace evaluado, siguiendo la aproximación GRADE[48].

El uso de este método en las RSL, propuesta por el grupo de trabajo "Grading of Recommendations Assessment, Development and Evaluation (GRADE Working Group"), busca evaluar "la certeza" que se puede tener a la hora de estimar el efecto de una intervención sobre los desenlaces evaluados; o sea, la probabilidad de que el efecto real sea sustancialmente diferente del estimado en el metaanálisis de la revisión[49]. El grupo de trabajo GRADE sugiere basar esta evaluación, máximo, en cinco o seis desenlaces, definidos desde el inicio de la revisión, por ser considerados críticos o importantes, tanto para el paciente como para los clínicos y para los demás tomadores de decisiones en salud[50].

Al momento de escribir este capítulo, el grupo de trabajo GRADE ha desarrollado aproximaciones para evaluar la calidad del cuerpo de evidencia en RSL de intervenciones y de precisión de pruebas diagnósticas, y trabaja en la aproximación para revisiones de otro tipo de estudios. Aquí se presentan los conceptos básicos para evaluar la evidencia en RSL de intervenciones; los lectores interesados en evaluar la evidencia de RSL de pruebas diagnósticas pueden consultar el manual generado por el grupo de trabajo GRADE para aplicar las aproximaciones GRADE[50].

El grupo de trabajo GRADE considera para los desenlaces de efectividad de las intervenciones que los clínicos controlados proveen la evidencia de más alta calidad, pero que la certeza en el cuerpo de evidencia proveniente de los ECA puede verse afectada por cinco aspectos[50]:

- Limitaciones en el diseño de los estudios incluidos, que pueden significar un alto riesgo de sesgo.
- Inconsistencia en los resultados de los estudios (presencia de heterogeneidad no explicada).
- La evidencia que proporcionan los estudios es indirecta (la pregunta que se aborda en la RSL es diferente de la que responden los estudios en cuanto a la población, la intervención, el comparador o los desenlaces evaluados).
- Imprecisión en los resultados (los estudios que aportan evidencia incluyen relativamente a pocos pacientes y pocos eventos, por lo cual la estimación agrupada del efecto está acompañada de IC amplios e imprecisos).
- Sesgo de publicación (la estimación del efecto puede estar sobreestimada o subestimada de forma sistemática, y ello se debe a la publicación selectiva de estudios).

La mayoría de los mencionados aspectos se han presentado anteriormente como parte de los pasos que se siguen en la conducción de la revisión; aquí les corresponde a los revisores juzgar de forma explícita el efecto que cada uno de ellos puede tener en la calidad global de la evidencia obtenida para cada desenlace. Ante la presencia de alguno de ellos (o de varios), la "certeza" en la evidencia para el desenlace en evaluación decrece. Por tanto, a mayor cantidad de problemas menor será la "certeza" en la evidencia hallada en la RSL[50].

Los revisores deben identificar en los hallazgos de la revisión los factores que llevan a reducir la calidad de la evidencia para cada desenlace en cada uno de los cinco aspectos mencionados, con el fin de identificar qué tan importante es la limitación del cuerpo global de la evidencia para cada aspecto en particular. Se acostumbra clasificarla como: "*no seria*", "*seria*" o "*muy seria*". Enseguida se presentan los factores para considerar en esta evaluación.

Factores que llevan a reducir la calidad de la evidencia que aportan los ECA

La **tabla 20.5** presenta los criterios para evaluar las limitaciones en el diseño de los estudios incluidos para cada desenlace. Se considera que hay limitaciones en el diseño cuando los estudios incluidos presentan:

- Falta de ocultamiento de la asignación.
- Falta de enmascaramiento (en particular, si los desenlaces son subjetivos y su evaluación es muy susceptible a sesgos).
- Pérdidas grandes durante el seguimiento.
- Falta de adherencia en el análisis al principio de intención de tratar.
- Suspensión temprana del reclutamiento, por identificar beneficio.
- Riesgo de reporte selectivo de eventos (por ejemplo, los investigadores omiten reportar desenlaces que han sido medidos; especialmente, los desenlaces para los cuales no encontraron efecto alguno.

Se considera que hay inconsistencia cuando los resultados que se basan en más de un estudio presentan[50]:

- Heterogeneidad no explicada en los resultados estimados del efecto (**RR**, **RRI**, *hazard ratio*).
- Amplia varianza de las estimaciones puntuales entre los estudios (nota: la dirección del efecto no es un criterio para la inconsistencia).
- Mínima superposición (o ninguna) de los **IC** en la presentación gráfica de los resultados **(figura 20.1)**, lo cual sugiere que la variación es más que lo esperable como fruto del azar.
- Bajo valor de p ($p < 0{,}05$), lo cual indica que se rechaza la hipótesis nula en los criterios estadísticos que incluyan pruebas de

Tabla 20.5 Criterios para evaluar las limitaciones en el diseño de los estudios incluidos para cada desenlace

Riesgo de sesgos	Entre estudios	Interpretación	Consideraciones	Valoración de las limitaciones del estudio
Bajo	La mayoría de la información proviene de estudios con bajo riesgo de sesgo.	Los sesgos plausibles tienen poca probabilidad de alterar seriamente los resultados.	Sin limitaciones aparentes.	Sin limitaciones serias.
No claro	La mayoría de la información proviene de estudios con riesgo de sesgo bajo o no claro.	Los sesgos plausibles generan algunas dudas sobre los resultados.	Las limitaciones potenciales tienen poca probabilidad de disminuir la confianza en el estimado del efecto.	Sin limitaciones serias.
			Las limitaciones potenciales tienen probabilidad de disminuir la confianza en el estimado del efecto.	Limitaciones serias.
Alto	La proporción de información que proviene de estudios con alto riesgo de sesgo es suficiente para afectar la interpretación de los resultados.	Los sesgos plausibles debilitan seriamente la confianza en los resultados.	Limitación crucial para un criterio, o algunas limitaciones para múltiples criterios, suficiente para disminuir la confianza en el estimado del efecto.	Limitaciones serias
			Limitación crucial para uno o más criterios, suficiente para disminuir sustancialmente la confianza en el estimado del efecto.	Limitaciones muy serias.

Fuente: Tomada de **GRADE Handbook. Disponible en: http://www.guidelinedevelopment.org/handbook/#h.xivvyiu1pr3v. Traducción** al español: María Ximena Rojas.

heterogeneidad (las cuales, a su vez, ponen a prueba la hipótesis nula de que todos los estudios tienen la misma magnitud subyacente del efecto).
- Resultado del estadístico I^2 **mayor que el** propuesto. Dicho estadístico cuantifica la proporción de la variación en las estimaciones puntuales, debido a diferencias entre-estudio.

Antes de considerar una reducción en la calidad de la evidencia por inconsistencia, los revisores deben evaluar la posibilidad de que esta se deba a diferencias en:
- Poblaciones (por ejemplo, algunos medicamentos pueden tener un efecto relativamente más grande en las poblaciones más enfermas).
- Intervenciones (por ejemplo, efectos más grandes con dosis más altas).
- Desenlaces (por ejemplo, disminución del efecto del tratamiento con el paso del tiempo).

En estos casos, es pertinente presentar la síntesis de evidencia por subgrupos de población, por subgrupos de intervención o según el método de evaluación que se empleó para medir los desenlaces, y evaluar la inconsistencia de los resultados en cada subgrupo[50].

Se considera que la evidencia es ***indirecta*** cuando:
- La población, la intervención, el comparador o el desenlace son indirectos para la pregunta que se trató de responder en la revisión. En otras palabras, la pregunta abordada por los autores de la revisión sistemática es diferente de la evidencia disponible en cuanto a la población, la intervención, el comparador o el desenlace.
- La comparación entre las intervenciones de interés de la revisión (A y B) no está disponible; es decir, no se identificaron estudios que compararan la intervención A con la intervención B, pero sí se identificaron estudios donde la intervención A fue comparada con una intervención C, y la intervención B también fue comparada con la intervención C. Dichos estudios permiten la comparación indirecta de la magnitud del efecto de A respecto a B, y, como parte de la revisión, los autores hicieron metaanálisis de comparaciones indirectas. Este tipo de evidencia es de menor calidad que la evidencia que aportaría una comparación cara a cara entre A y B.

Se considera que hay imprecisión cuando:
- Los resultados de la estimación del efecto para el desenlace de interés tienen IC **amplios alrededor del estimado del** efecto.
- El tamaño total de la muestra (acumulativo) es menor que el tamaño óptimo de información (**TOI**, comparable al cálculo de tamaño de muestra para un estudio).
- Al agrupar los estudios para la estimación agrupada del efecto, el número total de eventos (número de sujetos que presentan el desenlace) es menos de 300. Esa es una "regla de oro" que surgió a partir de simulaciones que han evaluado el efecto del riesgo basal y del tamaño del efecto[50].
- El **IC 95 % alrededor de la estimación agrupada** del efecto (**RR** o **OR**) incluye entre sus límites la posibilidad de no efecto de la intervención, así como la posibilidad de un efecto "benéfico apreciable" o un de "daño apreciable".

Por último, se considera que el cuerpo de evidencia para un desenlace está afectado por sesgo de publicación cuando:
- Los estudios que aportan evidencia para el desenlace en evaluación presentaron alto riesgo de reporte selectivo de desenlaces.

Al juzgar este criterio, los revisores deben considerar si existe:
- Publicación temprana de los resultados positivos.
- Estrategia de búsqueda integral y exhaustiva.
- Asimetría en el gráfico en embudo **(figura 20.2)**.

Es importante señalar que, si bien los factores que afectan la calidad de la evidencia en esta evaluación son aditivos, la definición del efecto de cada factor en la evidencia implica juicios que no son exclusivos, y, más bien, dependen de cómo se interprete la situación por parte de cada revisor. Debido a ello, quien hace la evaluación debe presentar una nota aclaratoria que permita entender por qué se aumentó o se redujo la calidad de la evidencia, y conocer

el juicio llevado a cabo en cada caso. De esta forma, el proceso cumplirá con la característica de ser explícito, transparente y reproducible[38]. Es altamente deseable que la evaluación de la calidad del cuerpo de evidencia obtenida para cada desenlace sea hecha al menos por dos revisores, de forma independiente, y que las diferencias en los resultados de la evaluación de cada aspecto sean discutidas hasta obtener un consenso[51].

Al final de este proceso, se tendrá el resultado sobre la "calidad de la evidencia", y, por tanto, sobre la certeza que se puede tener en las estimaciones del efecto de las intervenciones evaluadas. Aunque la calidad de la evidencia representa un continuo, el resultado de la evaluación siguiendo la aproximación GRADE puede caer en una de cuatro categorías: "alta", "moderada", "baja" o "muy baja". La **tabla 20.6** presenta el significado de cada una de esas categorías y sus implicaciones en la certeza que se puede tener en la evidencia o en los resultados finales de la RSL, como un insumo para la toma de decisiones[50].

EJEMPLO

En la evaluación de la calidad de la evidencia sobre el efecto del uso de probióticos profilácticos para prevenir la enterocolitis necrotizante en recién nacidos prematuros que requieren ser admitidos en unidades de cuidado neonatal, hay incertidumbre acerca de tres factores: 1) limitaciones en el diseño de algunos estudios, 2) inconsistencia de los resultados y 3) imprecisión de la estimación del efecto; sin embargo, dicha incertidumbre no es lo suficientemente grave como para degradar cada uno de ellos. Los revisores podrían, razonablemente, reducir la calidad de la evidencia o no hacerlo; un revisor puede, en cada aspecto, dar a los estudios el beneficio de la duda e interpretar la evidencia como de alta calidad, mientras que otro revisor podría descalificar la calidad de la evidencia en uno solo de los aspectos; por ejemplo "Seria imprecisión de la estimación del efecto", de tal forma que juzgaría la evidencia como de calidad moderada. Este último revisor debe proporcionar una nota tanto para los aspectos que decidió no descalificar como para el que decidió descalificar. En dichas notas, el revisor debe dejar claro el juicio que llevó a esas decisiones haciendo explícito que se había identificado la cierta incertidumbre (ver notas al pie de la **tabla 20.7**).

Tabla 20.6 Calificación de la certeza en la evidencia según la aproximación GRAD

Calificación	Definiciones	Implicaciones
Alta	La investigación provee un muy buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es muy baja.	La evidencia es una excelente base para la toma de decisiones sobre si implementar o no la intervención (uso de la tecnología). Es probable que no se requiera evaluación de impacto ni monitorear el impacto si la intervención es implementada.
Moderada	La investigación provee un buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es moderada.	La evidencia es una buena base para la toma de decisiones sobre si implementar o no una intervención. Es probable que se requiera monitoreo del impacto, y podría justificarse una evaluación del impacto si la intervención es implementada.
Baja	La investigación provee alguna indicación del efecto probable; sin embargo, la probabilidad de que el efecto sea sustancialmente diferente es alta.	La evidencia provee alguna base para la toma de decisiones sobre si implementar o no una intervención. Sin embargo, se justifica hacer una evaluación del impacto si la intervención es implementada.
Muy baja	La investigación no provee una indicación confiable sobre el efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es muy alta.	La evidencia no es una buena base para la toma de decisiones sobre si implementar o no una intervención. Preferiblemente, la intervención no debe ser implementada; si lo es, se requiere hacer una evaluación de su impacto.

Fuente: Tomada de GRADE Handbook, disponible en: http://www.guidelinedevelopment.org/handbook/#h.xivvyiu1pr3v. Traducción al español: María Ximena Rojas.

Tabla 20.7 Ejemplo de una tabla de resumen de hallazgos aplicada a RS de intervenciones

Probióticos comparados con placebo en recién nacidos prematuros (< 37 semanas, o peso < 2500 g)

Pacientes o población: **Recién nacidos prematuros (< 37 semanas, o peso < 2500 g)**
Intervención: Probióticos

Comparación: **Placebo**

Desenlaces	Riesgos comparativos* (IC 95 %)		Efecto relativo (IC 95 %)	No. de participantes (estudios)	Calidad de la evidencia (GRADE)
	Riesgo asumido	Riesgo correspondiente			
	Placebo	Probióticos			
Enterocolitis severa (estado II-III) Criterios de Bell Seguimiento: 1-6 meses.	**57 por 1000**	**25 por 1000** (19-32)	RR **0,43** (0,33-0,56)	**5529** (20 estudios)	⊕⊕⊕⊖ **moderada**[1]
Sepsis probada por cultivo (cualquier germen) Cultivo de sangre o de LCR tomado luego del quinto día de vida que es positivo. Seguimiento: 1-6 semanas.	**162 por 1000**	**148 por 1000** (130-167)	RR **0,91** (0,80-1,03)	**5338** (19 estudios)	⊕⊕⊕⊖ **moderada**[1]
Mortalidad (cualquier causa) Seguimiento: 1-6 semanas.	**70 por 1000**	**45 por 1000** (36-57)	RR **0,65** (0,52-0,81)	**5112** (17 estudios)	⊕⊕⊕⊖ **moderada**[2]
Tiempo para alimentación enteral plena Días para alimentación enteral plena. Seguimiento: 1-6 semanas.	El tiempo promedio en el grupo control fue de 14 **días**	El tiempo promedio en el grupo control fue **1,32 más corto (1,48-1,17** más corto)		2657 (8 estudios)	⊕⊕⊕⊖ **moderada**[3]

*La base para el riesgo asumido (e.g. la mediana del riego en el grupo control a través de los estudios) aparece en notas al pie. El riesgo correspondiente (y los IC 95 %) se basan en el riesgo asumido en el grupo control y el efecto relativo de la intervención (y sus **IC 95 %).**

IC: intervalo de confianza; RR: riesgo relativo; LCR: líquido cefalorraquídeo.

1 Los estudios de Al Hosni (2012), Bin-Nun (2005), Kitajima (1997), Manzoni (2006), Samanta (2009), Sari (2010) y Straiki (2007) no describen el ocultamiento de la asignación ni la evaluación ciega de los desenlaces. El seguimiento no fue completo en el estudio de Kitajima (1997).

2 Los estudios de Al Hosni (2012), Bin-Nun (2005), Kitajima (1997), Manzoni (2006), Samanta (2009), Sari (2010) y Straiki (2007) no describen el ocultamiento de la asignación ni la evaluación ciega de los desenlaces. El desenlace mortalidad no es afectado por esta limitación en el diseño. El seguimiento no fue completo en el estudio de Kitajima (1997) —razón para reducir el puntaje—.

3 Los estudios de Samanta (2009) y de Sari (2010) no describen el ocultamiento de la asignación ni la evaluación ciega de los desenlaces.

TABLAS DE RESUMEN DE HALLAZGOS

Para presentar la síntesis de la evidencia encontrada en la RSL para cada desenlace y la valoración que se hace de la calidad del cuerpo de evidencia, el grupo de trabajo GRADE propone hacer una tabla de resumen de hallazgos (*Summary of Findings table*, o SoFt, por sus siglas en inglés)[52]. Dicha tabla presenta de forma resumida los hallazgos finales de la RSL para cada uno de los desenlaces identificados como críticos e importantes (según la pregunta que la RSL buscó responder). Incluye el número de estudios que aportaron al cuerpo de evidencia para cada desenlace, el número de pacientes en los que se basan las estimaciones, los estimadores de efecto obtenidos como parte de los metaanálisis (si los hubo) y los IC que acompañan dichos estimadores, así como los resultados finales de la evaluación de la calidad de la evidencia[30].

Esta tabla se puede completar fácilmente a partir de los resultados de cada paso seguido en la revisión usando el paquete **GRADE pro** (disponible en: http://tech.cochrane.org/revman/other-resources/gradepro/download). El *Cochrane handbook for systematic reviews of interventions* **presenta un tutorial sobre cómo** completar estas tablas[30].

Los revisores deben evaluar, según la relevancia de sus hallazgos, la pertinencia de presentar una tabla SoFt para cada uno de los subgrupos estudiados o si la descripción narrativa de los resultados de la evaluación de la calidad del cuerpo de evidencia para los subgrupos en el texto de la revisión es lo bastante informativa para el lector. La **tabla 20.7 es un ejemplo de una** tabla SoFt en la que se presentan los resultados de la evaluación del cuerpo de evidencia obtenidos en la **RSL del uso de probióticos profilácticos** en recién nacidos prematuros.

CONCLUSIONES

Las **RSL son estudios de carácter integrativo** que pueden resumir, de forma válida y precisa, los resultados de múltiples estudios primarios.

Las **RSL, al igual que otros diseños de investigación** en salud, son susceptibles de errores aleatorios y sistemáticos. Para asegurar su validez, se hace necesario seguir de manera rigurosa procedimientos sistemáticos para la búsqueda, la selección, la evaluación y la síntesis de los estudios primarios, y que han sido descritos en este capítulo.

Los resultados finales de la evaluación y la síntesis del cuerpo de evidencia encontrada en la **RSL** para cada desenlace son la base para definir la certeza en las estimaciones de efecto obtenidas como parte de la **RSL, siguiendo la** aproximación **GRADE.**

El informe de los resultados de la **RSL debe** proporcionar al lector una descripción detallada de los métodos empleados, de manera que pueda evaluar críticamente los resultados de la revisión y tomar de ella los aportes a su situación particular.

Las **RSL de intervenciones, pruebas diagnósticas,** pronóstico y daño siguen la misma metodología; sin embargo, pueden presentar variaciones en la aplicación de algunos métodos y procedimientos, de acuerdo con el tipo de estudios que las componen.

Referencias

1. Baker R. The enigma of health: the art of healing in a scientific age. Med Hist. 1997;41:397-9.
2. Chalmers I, Altman DG. Systematic reviews. Londres: BMJ Publishing Group; 1995.
3. Chambers D, Wilson PM, Thompson CA, et al. Maximizing the impact of systematic reviews in health care decision making: a systematic scoping review of knowledge-translation resources. Milbank Q. 2011;89:131-56.
4. Littell JH, Corcoran J, Piyai V. Introduction. En: Littell JH, Corcoran J, Piyai V. Systematic reviews and meta-analysis. New York: Oxford University Press; 2008. pp. 1-28.
5. Egger M, Smith GD, O´Rourke K. Rationale, potentials and promise of systematic reviews. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 3-19.
6. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 The Cochrane Collaboration [internet]. Marzo de 2011. [citado 2015 mar. 15]. Disponible en: www.cochrane-handbook.org.
7. Egger M, Smith GD. Principles of and procedures for systematic reviews. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 23-42.
8. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. Inter J Epidemiol. 1999;28:1-9.
9. Egger M, Smith GD, Schneider M. Systematic reviews of observational studies. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 211-27.
10. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 248-82.
11. Norris SL, Atkins D, Bruening W, et al. Observational studies in systematic [corrected] reviews

of comparative effectiveness: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64:1178-86.
12. Reeves BC, Deeks JJ, Higgins JPT, et al. Including non-randomized studies. En: Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [internet]. Marzo de 2011. [citado 2015 mar. 15]. Disponible en: http://community.cochrane.org/handbook
13. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. J Clin Epidemiol. 2010;63:502-12.
14. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med. 2005;142(12 Pt 2):1090-9.
15. McClinton Appollis T, Lund C, de Vries PJ, et al. Adolescents' and adults' experiences of being surveyed about violence and abuse: a systematic review of harms, benefits, and regrets. Am J Public Health. 2015;105:e31-45.
16. Berg RC, Odgaard-Jensen J, Fretheim A, et al. An updated systematic review and meta-analysis of the obstetric consequences of female genital mutilation/cutting. Obstet Gynecol Int. 2014;2014:542859.
17. Thorat MA, Cuzick J. Prophylactic use of aspirin: systematic review of harms and approaches to mitigation in the general population. Eur J Epidemiol. 2014 Nov 25. [En prensa].
18. Cochrane Handbook DTA Reviews. The Cochrane Collaboration [internet]. 2013 [citado 2015 mar. 15]. Disponible en: http://srdta.cochrane.org/handbook-dta-reviews
19. Mulrow CD. Rationale for systematic reviews. En: Chalmers I, Altman D. Systematic reviews. Londres: BMJ Publishing Group; 1995. pp. 1-8.
20. Clarke M. Systematic reviews and the Cochrane Collaboration. Br J Surgery. 2007;4:391-2.
21. Egger M, Dickersin K, Smith GD. Problems and limitations in conducting systematic reviews. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 43-68.
22. Oxman AD, Cook DJ, Guyatt GH. User´s guides to the medical literature: VI. How to use an overview. JAMA. 1994;272:1367-71.
23. Riegelman RK. Meta-analysis. En: Riegelman RK. Studying a study and testing a test. How to read the Medical Evidence. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2000. pp. 85-199.
24. Hopayian K. The need for caution in interpreting high quality systematic reviews. BMJ. 2001;323:681-4.
25. Petticrew M. Why certain systematic reviews reach uncertain conclusions. BMJ. 2003;326:756-8.
26. The Question. En: National Health and Medical Research Council. How to review the evidence: systematic identification and review of the scientific literature. Commonwealth of Australia; 2000. pp. 9-14.
27. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med. 1997;127:380-7.
28. Khan KS, Kleijnen J. Selection of studies. En: Khan KS, Ter Riet G, Glanville J, et al. (eds.) for the NHS Centre for Review and Dissemination (CDR). Undertaking Systematic Reviews of Research on Effectiveness. CDR´s Guidance for Carrying Out or Commissioning Reviews. 2nd ed. CDR Report No. 4. York: NHS Centre for Reviews and Dissemination, University of York; 2000. pp. 33-6.
29. Clarke M, Oxman AD (eds). Locating and selecting studies. Cochrane Reviewers Handbook 4.1.3 [internet]. Junio de 2001. Section 5. The Cochrane Library, Issue 3, 2001. Oxford: Update Software.
30. Schünemann H, Oxman A, Higgins J, et al. Presenting results and 'Summary of findings' tables. En: Higgins JPT, Green S (eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [internet]. Marzo de 2011. [citado 2015 mar. 15]. The Cochrane Collaboration, 2011. Disponible en: www.cochrane-handbook.org.
31. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64:395-400.
32. Lefebvre C, Clarke MJ. Identifying randomised trials. En: Egger M, Davey Smith G, Altman DG. (eds.) Systematic Reviews in Health Care: Meta-analysis in context. 2ª ed. Londres: BMJ Publishing Group; 2001. pp. 69-86.
33. Petitti D. Information retrieval. En: Petitti D. Meta-Analysis, decision analysis and cost-effectiveness analysis. Oxford: Oxford University Press; 1994. pp. 43-62.
34. A search strategy for primary studies of effectiveness (Appendix 6). A general strategy for a more effective search of MEDLINE (Appendix 7). En: Jordan R. Cummins C, Burls A. (eds.) West Midlands development and evaluation service handbook. Version 2.2. 2000. pp. 111-5.
35. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. En: Chalmers I, Altman DG. Systematic reviews. Londres: BMJ Publishing Group; 1995. pp. 17-36.
36. Petitti DB. Advanced issues in meta-analysis. En:

Petitti DB. Meta-analysis, decision analysis and cost-effectiveness analysis. New York: Oxford University Press; 1994. pp. 70-84.
37. Juni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 87-105.
38. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273:408-12.
39. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609-13.
40. Petitti D. Data collection. En: Petitti D. Meta-Analysis, decision analysis and cost-effectiveness analysis. New York: Oxford University Press; 1994. pp. 63-9.
41. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. En: Higgins JPT, Green S (eds.). Cochrane handbook for systematic reviews of interventions version 5.1.0 [internet]. Marzo de 2011. The Cochrane Collaboration, 2011. Disponible en: http://community.cochrane.org/handbook
42. Petitti D. Statistical methods in meta-analysis. En: Petitti D. Meta-Analysis, decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine. New York: Oxford University Press; 1994. pp. 90-114.
43. Letón E, Pedromingo A. Introducción al análisis de datos en metaanálisis. Madrid: Ediciones Díaz de Santos SA; 2001.
44. Thompson SG. Why and how sources of heterogeneity should be investigated. En: Egger M, Davey Smith G, Altman DG. (eds.): Systematic reviews in health care: Meta-analysis in context. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 157-75.
45. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. En: Egger M, Davey Smith G, Altman DG. (eds.). Systematic reviews in health care: meta-analysis in context. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 285-312.
46. Russell K, Wiebe N, Saenz A, et al. Glucocorticoids for croup. Cochrane Database Syst Rev. 2004:CD001955.
47. Sterne JAC, Egger M, Smith GD. Investigating and dealing with publication and other biases. En: Egger M, Smith GD, Altman DG. Systematic reviews in health care. 2nd ed. Londres: BMJ Publishing Group; 2001. pp. 189-208.
48. Schünemann H, Oxman A, Higgins J, et al. Assessing the quality of a body of evidence. En: Higgins JPT, Green S (eds.). Cochrane handbook for systematic reviews of interventions version 5.1.0 [internet]. Marzo de 2011. [citado 2015 mar. 15]. The Cochrane Collaboration, 2011. Disponible en: www.cochrane-handbook.org.
49. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490-4.
50. GRADE Working Group. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Schünemann H, Brož ek J, Guyatt G, et al. (eds.). [Libro electrónico] Octubre 2013. Disponible en: http://www.guidelinedevelopment.org/handbook/
51. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. J Clin Epidemiol. 2013;66:736-42; quiz 742.e1-5.
52. Langendam MW, Akl EA, Dahm P, et al. Assessing and presenting summaries of evidence in Cochrane Reviews. Syst Rev. 2013;2:81.

CAPÍTULO

21

Introducción al análisis de decisiones

HOOVER QUITIÁN REYES

El ejercicio de las ciencias de la salud tiene como eje fundamental la toma de decisiones en todos los niveles: solicitar o no una resonancia magnética con contraste (un examen de alto costo), iniciar o no una terapia con efectividad demostrada, pero con riesgo significativo de efectos adversos, aconsejar o no una vacuna en personas mayores, entre otros, son ejemplos de la actividad diaria de un profesional de la salud. Siempre está tomando decisiones, y casi siempre, con bases que tienen un alto contenido de incertidumbre.

LA ELECCIÓN RACIONAL

Una decisión es, esencialmente, la escogencia, entre dos o más alternativas, de un curso de acción[1].

Casi siempre, con la existencia de, por lo menos, dos posibles formas de actuar mutuamente excluyentes, llamadas ***alternativas***, y la necesidad de determinar al menos un objetivo[2].

La elección resultante debería ser, aunque no siempre ocurre así, el resultado de un proceso racional en el que, a través de un proceso lógico, se determine, de acuerdo con la información disponible, cuál es la mejor de las alternativas.

El proceso de toma de decisiones de manera racional puede ser tan complejo que en ocasiones resulta más eficiente omitirlo[3-5].

Y actuar conforme a alguna regla preestablecida, que en muchos casos se originó de manera inconsciente[6]. A menudo, lo que el individuo considera que fue un proceso de elección racional no pasa de haber sido una acción basada en una regla autoimpuesta o en un hábito[7-9].

Como ejemplo de lo anterior, en un servicio de urgencias son muchas las decisiones que el personal médico debe tomar para salvaguardar la vida de los pacientes. Sin embargo, aunque se esperaría que todas sean producto de un proceso racional riguroso, debido al tiempo tan corto del cual se dispone para tomar la decisión, probablemente se actúe de manera instintiva, empleando de manera eficiente información de eventos pasados o siguiendo un protocolo preestablecido.

A diferencia de lo que ocurre en las acciones de emergencia, la mayoría de las decisiones que debe tomar el personal médico son resultado de

un proceso racional que incorpora resultados de estudios clínicos enriquecidos mediante la consulta a expertos y la experiencia personal (medicina basada en evidencia) y los considera a la luz de las preferencias del paciente[10].

Enseguida se exponen los elementos básicos de la ***teoría de la decisión***, que se ocupa de los procesos de toma de decisiones de manera racional. Se entiende por ***racionalidad*** el cumplimiento de ciertos axiomas por parte del individuo y la forma como este emplea la información de la cual dispone, así como el seguimiento de ciertos pasos que buscan garantizar que la elección hecha sea la mejor, dado el conjunto de información disponible. Se dice que un individuo es racional si sus preferencias (la manera como hace sus elecciones) son, al menos[11,12]:

- Completas: Todas las opciones pueden compararse en duplas, y así determinarse cuál es preferida sobre la otra, o si son preferidas por igual.
- Reflexivas: Cada alternativa es al menos tan preferida como ella misma u otra opción igual, en todas sus dimensiones y sus características.
- Transitivas: Si se decide que la opción A es preferida a la opción B, y que la opción B es, a su vez, preferida sobre C, entonces puede afirmarse que A debe ser preferida a C.

ELEMENTOS DEL PROBLEMA DE DECISIÓN

Para que se pueda hallar solución válida a un problema de decisión, se debe contar con los siguientes elementos:

- Un decisor racional: encargado de seleccionar qué alternativa seguir.
- Las alternativas o acciones: todas las posibles formas de actuar.
- Estados de la naturaleza: todos los eventos o circunstancias presentes o futuras que escapan del control del decisor e influyen en el resultado.
- Las consecuencias: resultados obtenidos con cada una de las alternativas según cada uno de los posibles estados de la naturaleza.
- Una regla de decisión: procedimiento para identificar la mejor de las alternativas empleando como insumo la valoración de sus resultados.

LAS FASES DEL PROCESO DE DECISIÓN

Todo proceso de elección racional debe seguir los siguientes pasos para asegurar que la decisión final sea la mejor posible:

- Identificación de las alternativas: De manera exhaustiva, deben identificarse todas las alternativas de acción posibles; de lo contrario, existe el riesgo de que el proceso de decisión no sea válido. En vano se adelantarán los demás componentes del proceso de decisión si hay alternativas posibles que no fueron evaluadas, por lo cual no podrá asegurarse que el resultado corresponda con el mejor accionar.
- Predicción de los resultados: Cada una de las posibles acciones tendrá consecuencias asociadas, que pueden ocurrir de manera inmediata o en el futuro. Estas deberán ser descritas de manera completa, y todas ellas, agotar los eventos que puedan ocurrir. En caso de que no se establezcan los resultados de cada alternativa, o de que estos estén caracterizados de manera parcial, la comparación entre ellos no podrá efectuarse adecuadamente, y se corre el riesgo de llegar a conclusiones incorrectas. No basta con enunciar los eventos: también es necesario que, si es posible, se consulten o sean calculadas las probabilidades de ocurrencia. Aquí, la revisión sistemática de la literatura, los metaanálisis y la consulta a expertos constituyen fuentes de información.
- Valoración de los resultados: Los resultados de las acciones posibles deben compararse a la luz de los objetivos que se busca alcanzar. Para ello, es necesario asignar un valor o un símbolo a cada uno de ellos, que permita, como mínimo, un ordenamiento.

Elección

Usando las valoraciones, se hace la elección de acuerdo con un criterio de decisión que permita el ordenamiento de las alternativas. Dependiendo del conjunto de información, el criterio de decisión puede no ser fácil de determinar de manera totalmente satisfactoria. Sin importar cuál sea, debe garantizar el cumplimiento de ciertos axiomas, para que el proceso de decisión se considere racional. Los siguientes son los axiomas de la racionalidad sobre los criterios de decisión:

Orden completo

El criterio de asignación debe permitir un ordenamiento de todos los resultados, sin haber alternativas que no sean comparables frente a las demás.

Simetría

El criterio de decisión no debe depender del orden en el que se enumeran las alternativas o los resultados.

Linealidad

El orden no debe cambiar si los valores son reemplazados por transformaciones lineales de ellos mismos; es decir, si todos son multiplicados por la misma constante > 0, o si a todos se les suma el mismo valor, el ordenamiento deberá permanecer inalterado.

Independencia de alternativas irrelevantes

El orden entre dos alternativas no debe cambiar si se adiciona una nueva alternativa.

Independencia de permutación de alternativas

Si existen dos alternativas con los mismos resultados, estos deberán contar con la misma valoración.

REPRESENTACIÓN

Las formas más comunes como se representan los problemas de decisión son las ***tablas*** y los *árboles*; ambos permiten visualizar completo el conjunto de información.

En general, las tablas de decisión ubican en las ***columnas*** los estados de la naturaleza, y en las *filas*, las alternativas; ello, por cuanto es muy probable que sean más las alternativas que los estados, y se prefiere contar con más filas que columnas. La primera fila se destinará a los encabezados de los estados, y la primera columna, a la enumeración de las alternativas; de esta manera, cada una de las casillas ubicadas en filas y columnas de orden mayor de dos deberán contener el resultado o la respectiva valoración:

	Estado 1	Estado 2	…	Estado *n*
Alternativa 1				
Alternativa 2				
Alternativa 3				
…				
Alternativa *n*				

Los árboles de decisión, por su parte, permiten visualizar con mayor claridad la cadena de eventos que pueden ocurrir de manera simultánea o en secuencia. En el árbol de decisión, las alternativas son conocidas como ***nodos de decisión*** (identificados con un ***cuadrado***), y los estados, como *eventos probabilísticos* (identificados con *círculos*). Cada nodo de decisión tendrá una rama independiente, que deberá seguir el curso de los posibles eventos, y así generar bifurcaciones, cada una de las cuales deberá llegar a un ***nodo terminal*** (identificado con un ***triángulo***) que dará fin a la cadena. Este nodo terminal será el equivalente de una de las casillas de orden > 2 que haya en la representación de tabla.



Una representación de árbol deberá tener tantos nodos terminales como casillas existan en la tabla de decisión.

Como el uso de árboles permite una diagramación mucho menos rígida que la de las tablas, se recomienda seguir estos lineamientos para evitar que la manera como se resuma la información pueda omitir alguno de los elementos relevantes[13]. En la construcción de árboles deben tenerse en cuenta varias recomendaciones:

Nodo de decisión solo al inicio

Para disminuir la complejidad del problema, se recomienda el uso de un árbol por cada decisión, y que esta solo aparezca en el nodo inicial.

Simetría del árbol

El estado inicial debe afectar a todos los resultados.

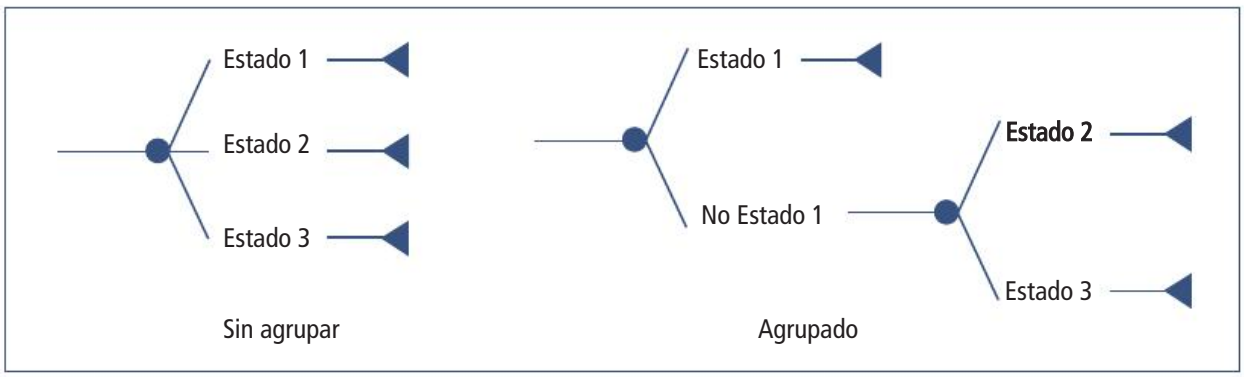
Balance del árbol

Es necesario evaluar la posibilidad de ocurrencia de todos los eventos para cada una de las alternativas. Por ejemplo, en el caso de las decisiones médicas, las intervenciones tienen eventos adversos y beneficios asociados. Un árbol que compare dos alternativas de tratamiento

y que solo muestre los eventos adversos de una de ellas, no estaría balanceado.

Nodos probabilísticos que tengan solo dos ramas

Si existen tres o más eventos posibles, para poder evaluar el cambio en la probabilidad de ocurrencia de cada uno de ellos, lo mejor es hacer agrupamientos de tal forma que solo existan dos brazos y que se pueda definir un único parámetro de probabilidad, y el otro, como su complemento (1-probabilidad del otro evento):



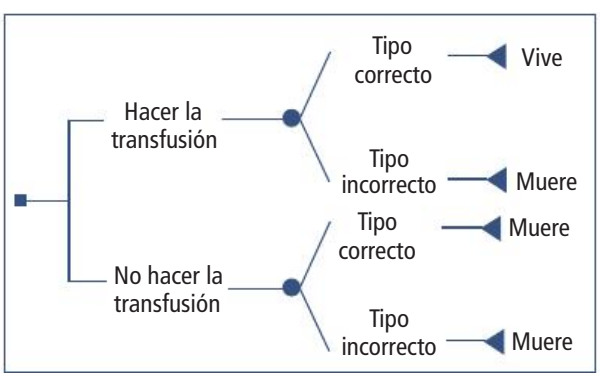
EJEMPLO

La condición de salud de un paciente requiere que se le haga una transfusión de sangre, y hasta el momento tan solo se sabe que existen diferentes tipos de sangre, pero no se ha determinado cuántos de estos hay ni cuáles son. Se sabe que si no se hace la transfusión, el paciente morirá, y que vivirá solo si se lo trata con el tipo adecuado.

Empleando la tabla de decisión:

		Estados de la naturaleza	
		Tipo de sangre	
		Correcto	Incorrecto
Alternativas	Hacer la transfusión	Vive	Muere
	No hacer la transfusión	Muere	Muere

En la representación de árbol:



En el ejemplo, las alternativas solo tienen dos resultados posibles: el paciente ***Vive*** o ***Muere***. Para poder emplear un criterio de decisión, es necesario asignar valores que reflejen las preferencias; es decir, que se prefiere Vive sobre Muere.

CLASIFICACIÓN DE LOS PROBLEMAS DE DECISIÓN

La clasificación se hace según el conjunto de información y la cantidad de variables que están bajo control de quien debe tomar la decisión.

Los procesos de certidumbre se caracterizan porque todos los actores son conocidos, así como las posibles acciones de cada uno y sus respectivas consecuencias. El ejemplo anterior podría considerarse de certeza solo si se conociera el tipo de sangre del paciente y el tipo de sangre disponible para la transfusión.

La existencia de riesgo, o ***incertidumbre***, se refiere a que las consecuencias de cada una de las acciones tiene asignada una probabilidad de ocurrencia. En el ejemplo de la transfusión, se hablaría de un escenario de riesgo si aunque no se tenga certeza del tipo de sangre del paciente, sí estén identificados todos los tipos de sangre posible y la probabilidad de que el paciente tenga cada uno de ellos.

La incertidumbre total se refiere a la imposibilidad de asignar probabilidades de ocurrencia

de los posibles eventos, ya sea porque son estas son desconocidas o porque no tiene sentido definirlas. Es el caso del ejemplo inicial, pues no se conocen los tipos de sangre existentes, y por tal razón no es posible asignar distribución de probabilidad alguna.

CRITERIOS DE DECISIÓN SEGÚN EL TIPO DE PROBLEMA

Dependiendo del tipo de problema, el criterio de decisión puede ser único o habrá varias alternativas que cumplan con los axiomas de racionalidad.

Criterio de elección en un problema con certidumbre

Si el problema de decisión está en un escenario de certidumbre, el hecho de que se sepa con certeza cuáles serán los eventos que ocurrirán y sus respectivos resultados hace que exista un único criterio de decisión, el cual consiste en ordenar los resultados según su respectiva valoración, y se elige la alternativa que signifique el mayor valor.

Criterios de elección en problemas con riesgo o bajo incertidumbre

En los problemas donde, dada la elección hecha, solo se conoce la probabilidad de ocurrencia de cada uno de los eventos, hay varias alternativas. Las siguientes son algunas de las más utilizadas:

El valor esperado

Para cada una de las alternativas se calcula el valor esperado como la suma ponderada de todos sus posibles resultados, y empleando como ponderación la probabilidad de ocurrencia. Se escogerá la alternativa con el mayor valor esperado.

Varianza mínima y media acotada

Cada valor esperado tiene asociada una varianza. Este criterio consiste en determinar un nivel mínimo que se espera obtener como resultado de la acción; de esta manera, se considerarán únicamente las alternativas que, al menos, tengan ese valor esperado, y de estas se escogerá la que cuente con la menor varianza.

Media con varianza acotada

En este criterio se determina el máximo de varianza que se estaría dispuesto a aceptar, para luego escoger, entre las alternativas que reporten ese nivel de varianza, o menor, la que tenga el mayor valor esperado.

Criterios de elección bajo incertidumbre total

En los casos donde no es posible determinar la probabilidad de ocurrencia de los eventos, pueden emplearse criterios que solo utilicen el ordenamiento de las valoraciones de los resultados:

Maximin, o criterio de Wald[14]

Debido a que cada alternativa podrá tener diferentes desenlaces o resultados (cada uno de ellos, a su vez, con un valor numérico asignado), una regla de decisión puede consistir en comparar los valores mínimos y escoger la alternativa con el mayor valor entre los mínimos. A dicha estrategia se la considera un pensamiento pesimista, pues asume que siempre se dará el menos favorable de los escenarios.

Maximax

Similar a la anterior, en este caso se comparan las alternativas empleando los valores máximos de los resultados, por lo que se la considera un criterio de pensamiento optimista.

Criterio de Laplace

Como no se tiene información sobre la distribución de probabilidad, se asume que cada evento tiene la misma probabilidad de ocurrencia para calcular el valor esperado.

CONCLUSIONES

Para tomar decisiones es necesario haber hecho un cuidadoso análisis de los potenciales daños, así como de los costos y los beneficios asociados a las diferentes opciones de conducta. Generalmente, las decisiones se toman en presencia de prioridades que compiten, de recursos

limitados y de información incompleta o con una alta dosis de incertidumbre.

En esas condiciones, resulta fundamental el ejercicio, riguroso y tan objetivo como sea posible, del análisis de los desenlaces y de sus probabilidades, para ofrecer información que aporte al proceso de decisiones.

Una herramienta muy útil para el análisis de decisiones es el árbol de decisiones, que representa las opciones con sus probabilidades y sus consecuencias. Los resultados ofrecen insumos valiosos al tomador de decisiones, y pueden permitir la toma de decisiones clínicas con mejor información.

Referencias

1. Bonome MG. La racionalidad en la toma de decisiones: análisis de la teoría de la decisión de Herbert A. Simon. Oleiros, La Coruña: Netbiblo; 2009.
2. Peterson M. An introduction to decision theory. New York: Cambridge University Press; 2009.
3. Gazzaniga MS. Who's in charge?: free will and the science of the brain. New York: Ecco; 2012.
4. Fuster JM. The prefrontal cortex-An update-Time is of the essence. Neuron. 2001;30:319-33.
5. Iyengar S. The art of choosing. New York: Twelve; 2011.
6. Akerlof GA, Shiller RJ. Animal spirits: how human psychology drives the economy, and why it matters for global capitalism. Princeton: Princeton University Press; 2009.
7. Duhigg C. The power of habit: why we do what we do in life and business. New York: Random House; 2012. p. 371.
8. Hodgson GM. La ubicuidad de los hábitos y las reglas. Rev Econ Inst. 2000;2:11-43.
9. Bazerman MH, Neale MA. La negociación racional: en un mundo irracional. Barcelona: Paidós; 1993.
10. Detsky AS, Naglie G, Krahn MD, et al. Primer on medical decision analysis: Part 1-Getting started. Med Decis Mak Int J Soc Med Decis Mak. 1997;17:123-5.
11. Nicholson W. Teoría microeconómica: Principios básicos y ampliaciones [libro electrónico]. Cengage Learning Latin America; 2006. [citado 2015 feb. 9]. Disponible en: http://books.google.com.co/books?id=D-tRgvKyrikC
12. Varian HR. Microeconomía intermedia: Un enfoque actual [libro electrónico]. Bosch A. (Ed.). 2003. [citado 2015 feb. 9]. Disponible en: http://books.google.com.co/books?id=-rWfisFu618C
13. Gehrke J. Decision tree. En: Nong Y. The handbook of data mining. New Jersey: Lawrence Erlbaum Associates; 2003. pp. 3-23.
14. Wald A. Statistical decision functions. New York: Wiley; 1950.

CAPÍTULO

22

Desarrollo de guías de práctica clínica

Óscar Mauricio Muñoz Velandia • Ana María de la Hoz Bradford

INTRODUCCIÓN

Durante las últimas décadas se ha buscado cada vez más la incorporación del conocimiento científico en la práctica de la medicina. En respuesta a ello, la cantidad de investigación clínica se ha incrementado de forma explosiva, lo que conlleva que los médicos tengan a su disposición cantidades enormes de información. Resulta, pues, muy difícil para los profesionales de la salud mantenerse actualizados, conocer la calidad de la evidencia disponible e incorporarla en su práctica diaria.

Las guías de práctica clínica (GPC) tienen como objetivo no solo sintetizar esta clase de información, sino generar recomendaciones respecto a lo que se debe y no se debe hacer en el manejo de una determinada condición de salud, con el fin de mejorar los estándares de atención identificando las estrategias más adecuadas (o sea, seguras y efectivas). Cuando están bien desarrolladas, las GPC facilitan la labor tanto del personal de salud como la de los administradores y los políticos del sector, y eso, a su vez, redunda en beneficios para los pacientes y sus familias.

DEFINICIÓN, VENTAJAS Y DESVENTAJAS DE LAS GUÍAS DE PRÁCTICA CLÍNICA

Las GPC son definidas por el Instituto de Medicina de los Estados Unidos como "el conjunto de recomendaciones desarrolladas de manera sistemática, para ayudar a los clínicos y a los pacientes en el proceso de la toma de decisiones, sobre cuáles son las intervenciones más adecuadas para resolver un problema clínico en circunstancias sanitarias específicas"[1].

Como puede verse, el énfasis fundamental de las guías está en apoyar la toma de decisiones, lo cual las hace diferentes de otras formas de investigación integrativa en salud.

Se han delimitado algunas situaciones donde la toma de decisiones es muy difícil, por lo que es recomendable desarrollar y tener una GPC[2-5]:

- Cuando existe una gran variabilidad o incertidumbre en el abordaje de alguna entidad.
- Cuando no hay consenso para manejar a los pacientes con una condición.

- Cuando haya problemas o situaciones clínicas de alto impacto social o económico.
- Cuando una adecuada práctica pueda reducir significativamente la morbilidad y la mortalidad producidas por una enfermedad.
- Cuando las pruebas diagnósticas o los tratamientos producen efectos adversos o costos innecesarios.

En todas las situaciones mencionadas, el desarrollo de GPC puede producir múltiples beneficios. El principal, la contribución a mejorar la efectividad del cuidado clínico y de la calidad de la atención. Además, las GPC pueden ayudar a contener el costo y a racionalizar la demanda, progresivamente creciente, de los servicios de salud, a través de recomendaciones que definan el uso correcto de las tecnologías existentes; tanto las diagnósticas como las terapéuticas. Así mismo, pueden facilitar la generación de políticas de salud que incrementen la disponibilidad y faciliten el acceso a tecnologías que claramente se traduzcan en beneficios para los pacientes.

También, sin embargo, existen múltiples voces en contra del desarrollo de GPC[6]. Quienes se oponen a las GPC argumentan que estas no han demostrado cambios significativos en los procesos de salud ni, menos aún, en los desenlaces finales de los pacientes. En una revisión de 59 evaluaciones de GPC publicadas, solo en 4 de ellas se documentaron mejorías significativas en los procesos de atención, y solo en 2 casos se observaron impactos de alguna clase en los desenlaces en salud[7].

Lo anterior nos muestra cómo las GPC no son herramientas que funcionen por sí solas, y deben estar acompañadas por otros esfuerzos útiles para cambiar el comportamiento clínico. Las GPC deberán, pues, articularse con procesos de implementación, y deberán ser diseñadas por los entes gestores de las políticas en salud y apoyadas por las sociedades científicas y por los entes académicos.

Otras críticas argumentan que las recomendaciones formuladas por las GPC no toman en cuenta las características únicas de cada paciente, dada la alta frecuencia de comorbilidades y el hecho de que ninguna GPC puede hacer recomendaciones específicas para cada caso particular[6]. Así mismo, opinan que las GPC limitan la probabilidad de tomar decisiones consensuadas entre los médicos y los pacientes, al no tener en cuenta las preferencias de estos últimos[8]. Finalmente, algunas encuestas aplicadas entre médicos internistas demuestran resistencia a la implementación y el uso de estas guías, pues sienten que las GPC son sobresimplificadas y rígidas, y limitan la autonomía médica[9].

Cabe aclarar, sin embargo, que las GPC no tienen como intención ser camisas de fuerza, con protocolos estrictos y rígidos, sino que, por el contrario, ofrecen un direccionamiento general, sugieren unos estándares de cuidado que permitan tener maniobrabilidad teniendo en cuenta las características individuales de los pacientes, sus preferencias, la severidad de la condición clínica y la disponibilidad de los tratamientos en el ambiente donde se esté desarrollando la práctica. Resulta claro, pues, que las GPC no deberán ser usadas con fines de penalizar (multar) a los médicos, ni para definir si una conducta es ética o no[2].

Como una reflexión importante, una GPC debe ser *el comienzo* de un proceso de toma de decisiones, consensuadas con el paciente, y no el final de una decisión. Esto enfatiza la importancia de tener una base de discusión (la GPC), que, a la luz de las condiciones específicas del paciente individualizado y de sus preferencias, permita tomar una decisión. Como puede concluirse, la GPC es el primer paso, es una orientación, y no una camisa de fuerza. Y la GPC no puede por sí misma cambiar la práctica: también es necesaria su implementación correcta.

MÉTODOS PARA DESARROLLAR GPC

Existen diferentes métodos para el desarrollo de las GPC. Los más utilizados a través del tiempo han sido el consenso informal, el consenso formal y las GPC basadas en evidencia[10].

Consenso informal

El consenso informal es el método más antiguo de desarrollo de GPC. En dicha estrategia, un grupo de participantes (usualmente, expertos en el tema de interés), alcanzan consenso mediante discusión, y deciden qué se debe recomendar. Este fue un método de uso común, fácil y rápido de desarrollar, dado que no conlleva procedimientos analíticos complejos.

Su principal limitación radica en que no es posible evaluar la calidad de las recomendaciones, por la ausencia de un proceso de recolección sistemática y exhaustiva de la información; adicionalmente, no se presentan las bases que se tomaron en cuenta a la hora de generar las recomendaciones. De esta forma, las **GPC** desarrolladas por consenso informal pueden ser influidas (e, incluso, sesgadas) por personalidades dominantes o por consideraciones políticas, económicas o administrativas inapropiadas, lo que, en últimas, afecta la credibilidad del resultado.

Consenso formal

En dicha estrategia, un panel de expertos llega a un consenso mediante una serie de reuniones y conferencias que se hacen a puerta cerrada, durante dos días y medio, siguiendo un formato estructurado. El tercer día se presentan los resultados a una audiencia mayor y se hace una conferencia de prensa. Los miembros del equipo deben trabajar las horas que sean necesarias, dentro del tiempo establecido, para llegar a un acuerdo.

Si bien presenta avances respecto al método anterior, las limitaciones persisten en el consenso formal, pues aún no hay una metodología definida capaz de garantizar que la evidencia que fundamenta el consenso sea completa, y de buena calidad, y que no existen criterios explícitos para justificar las decisiones.

GPC basadas en evidencia

Buscando corregir las limitaciones de los métodos de consenso formal e informal, se construyó la metodología para elaborar **GPC** basadas en evidencia.

El Instituto de Medicina de los Estados Unidos publicó en 2010 el artículo "Guías de práctica clínica en las que podemos confiar"[11]. De acuerdo con dicha publicación, "las guías de práctica clínica son documentos que incluyen recomendaciones cuya intención es optimizar el cuidado del paciente, que están basadas en una revisión sistemática de la evidencia y que presentan un balance entre los beneficios y los riesgos de cada una de las alternativas de cuidado". La anterior definición permite hacer énfasis en las características de una **GPC basada** en evidencia, y las cuales mejoran la confianza que se tiene en sus recomendaciones finales:

1. **Una GPC basada en evidencia se desarrolla** teniendo en cuenta toda la información disponible, y seleccionando la evidencia con la mayor calidad posible. Para garantizar esto, las guías de buena calidad de esta clase deben demostrar que se hizo una revisión de la literatura de forma sistemática, que se presentan y se cumplen criterios explícitos para la inclusión o la exclusión de estudios y que se evaluó de manera estructurada la calidad de la información. Este proceso permite, adicionalmente, relacionar las recomendaciones con la calidad de la evidencia en la que se basan.
2. **Para ser confiable, una GPC basada en** evidencia debe presentar una metodología estructurada para generar recomendaciones. En este proceso se debe evaluar con cuidado el balance entre los beneficios y los riesgos de cada intervención. En lo posible, deben siempre incorporarse otras variables, tales como las preferencias de los pacientes y las consideraciones de costos.

Los documentos finales de las **GPC** desarrolladas de esta manera deberán incorporar de forma explícita toda la información que se consideró y las discusiones hechas a la luz de esta, en el proceso de generación de recomendaciones. Estos documentos deben ser evaluados por expertos externos, buscando incorporar en tal proceso opiniones de clínicos en ejercicio y de las sociedades científicas de todas las especialidades relacionadas con la condición a la que hace referencia la **GPC.**



Las **GPC** basadas en evidencia son, por todas las consideraciones mencionadas, las más aceptadas hoy día.

Sus inconvenientes radican en que a menudo no es posible seguir todos los pasos; ello ocurre, precisamente, debido a la falta de evidencia. Por esta razón, pueden encontrarse más o menos a menudo guías que hacen una mezcla entre una base de evidencia y el consenso de expertos. Una segunda limitante está dada porque resultan costosas y su desarrollo es demorado.

PASOS PARA DESARROLLAR GPC BASADAS EN EVIDENCIA

Selección del foco de una GPC

Las recomendaciones basadas en evidencia pretenden guiar al personal clínico y a los pacientes en los distintos procesos de toma de decisiones en el cuidado de la salud, dentro del marco de condiciones o enfermedades específicas. Es primordial la selección de temas prioritarios que requieran, o que puedan beneficiarse del desarrollo de GPC basadas en evidencia, considerando que su elaboración consume una gran cantidad de recursos (económicos y humanos) y de tiempo.

Actualmente no hay consenso respecto a cuál es el método más adecuado para priorizar los temas a los que se deben asignar recursos para investigación en salud; tampoco, para la priorización de condiciones a la hora de elaborar GPC. Se observa, además, una gran variabilidad en los métodos utilizados por diferentes organizaciones desarrolladoras de guías en este aspecto: en una encuesta internacional, solo el 30 % de las organizaciones reportó el uso de métodos explícitos o sistemáticos en la priorización de temas para elaborar guías. El resto de las organizaciones llevan a cabo procesos de priorización informales, tales como consultas a expertos y a actores interesados o métodos de consenso no formal[12].

Cabe anotarlo, se espera que los procesos de priorización sigan diferentes lineamientos y tengan niveles apropiados de objetividad y rigurosidad, dependiendo del contexto donde se lleven a cabo. En algunos casos, ciertas entidades privadas (instituciones prestadoras de servicios de salud —IPS—, centros médicos, aseguradoras) pueden asignar sus recursos siguiendo procesos de priorización internos que pueden no ser explícitos o transparentes para los pacientes ni para la población, pues siguen y sirven a los intereses propios de la entidad. Por el contrario, los entes gubernamentales y las organizaciones públicas (secretarías, ministerios de salud) deben seguir procesos de priorización explícitos, transparentes y sistemáticos que sirvan a los intereses de salud de la población general, con el fin maximizar el beneficio de la inversión de recursos públicos, que son siempre finitos[13].

En la literatura pueden hallarse publicados o recomendados diferentes métodos cualitativos y cuantitativos para priorizar temas que podrían beneficiarse del desarrollo de recomendaciones basadas en evidencia o GPC[14]; y si bien los criterios propuestos varían ampliamente, la mayoría de ellos dirige el proceso a identificar condiciones o áreas de salud en las que se identifique una carga de enfermedad importante para la población, incertidumbre sobre la efectividad de las alternativas para lograr los desenlaces deseados, o sobre el impacto potencial de las recomendaciones en la práctica y en la salud de la población, y la posibilidad de implementar las recomendaciones. La tabla 22.1 presenta un resumen de los criterios que suelen utilizarse con el fin de identificar prioridades a escala local, regional o nacional para la elaboración de GPC.

La aplicación de los criterios puede dirigir el proceso hacia la selección de temas relevantes para la población, y en los que el desarrollo de recomendaciones basadas en evidencia tenga un impacto importante. Adicionalmente, el proceso de priorización debe ser objetivo y transparente, y debe contar con la participación de trazadores de políticas públicas en salud, organizaciones académicas, asociaciones científicas, pacientes y la población general (todos los actores interesados en el proceso), con el fin de dar mayor validez y credibilidad a la toma de decisiones en la asignación de recursos para elaborar GPC[15,16]. Una vez se ha definido el foco de la guía (es decir, el

Tabla 22.1 Criterios para identificar prioridades en salud a escala local, regional o nacional para elaborar GPC

- Problemas en salud asociados a una carga de enfermedad importante para la población; enfermedades emergentes de alta relevancia.
- Ausencia de guías o recomendaciones basadas en evidencia válidas.
- Presencia de variabilidad no explicada en la práctica clínica, o diferencias entre la práctica actual y el cuidado apropiado; intervenciones ampliamente utilizadas, pero con efectividad cuestionable.
- La factibilidad para el desarrollo de recomendaciones que tengan impacto en desenlaces en salud, reducción de inequidades en salud o disminución de costos o de gastos para el sistema de salud con su implementación.
- Posibilidad para la implementación. No deben invertirse tiempo ni recursos en recomendaciones imposibles de implementar por la ausencia de recursos o por la presencia de barreras insuperables para su aplicación. Las recomendaciones deben ser aceptables para los usuarios potenciales.

Fuente: Adaptado de Oxman[14].

tópico y los objetivos para cumplir), se procederá a constituir el grupo que la desarrollará.

Constitución del grupo

La elaboración de una GPC involucra una gran cantidad de actividades y procesos que requieren conocimientos y experiencias en diversas áreas metodológicas y clínicas.

Y el grupo desarrollador de la guía debe estar conformado según los requerimientos de dichos procesos. Constituir equipos plurales y multidisciplinarios facilitará la formulación adecuada de recomendaciones, pues:

a. **Permite que el conocimiento requerido en** diferentes áreas se encuentre adecuadamente cubierto.
b. **Facilita la identificación del universo completo** de la evidencia científica.
c. **Aumenta el aval y la credibilidad de la guía,** al permitir una participación incluyente.
d. **Provee un espectro más amplio a la hora** de identificar potenciales barreras y facilitadores para implementar las recomendaciones[14,17].

El número de integrantes de un grupo desarrollador de una guía varía ampliamente, y depende, en gran medida, de los recursos disponibles y de la extensión o la complejidad de la guía en desarrollo[16]; la cantidad de aspectos abordados por una GPC **(por ejemplo, recomendaciones para el diagnóstico de la condición *vs*. abordaje completo de la enfermedad desde** el diagnóstico hasta el seguimiento y la rehabilitación) influirá de manera importante en la decisión sobre el número de personas que conformen el grupo, así como sobre la complejidad esperada de las labores, la cantidad y la calidad de la evidencia de la cual se anticipe que estará disponible.

Pueden distinguirse en el grupo desarrollador dos equipos con labores diferentes, pero complementarias:

a. **El *equipo técnico*, que constituye el principal** grupo de trabajo, y el cual se encarga de los procesos y de las actividades técnicas, metodológicas, logísticas y de consolidación de los informes, del reporte y la publicación del documento de la guía y de las recomendaciones.
b. **El *equipo de trabajo ampliado*, que participa,** principalmente, en labores específicas de apoyo técnico, consulta, consenso y votación[16].

La **tabla 22.2 describe la conformación convencional** de un grupo desarrollador.

Cabe resaltar la importancia de la incorporación de pacientes o de sus representantes a los equipos técnicos y de apoyo. Su participación en el proceso de desarrollo de la guía y la incorporación de sus puntos de vista y sus perspectivas en la toma de decisiones, a través de los diferentes procesos propios de la elaboración de la guía, posibilita generar recomendaciones más apropiadas y de mayor aplicabilidad[18-20]. Sin embargo, es importante tomar en cuenta que,

Tabla 22.2 Conformación típica del grupo desarrollador de una guía de práctica clínica

Organización	Descripción
Grupo desarrollador	
Equipo técnico	• Líder del grupo. • Expertos técnicos o metodológicos. • Profesionales clínicos (expertos en la condición de interés de la GPC). • Pacientes o cuidadores. • Expertos en economía de la salud.
Equipo de apoyo	• Colaboradores expertos (diferentes profesionales de la salud expertos en el área clínica). • Estadísticos. • Comunicadores. • Bibliotecarios. • Pacientes o cuidadores.

Fuente: Elaborado a partir de Carrasquilla[16].

pese a ser altamente recomendable, la participación de pacientes o de sus representantes como miembros del grupo desarrollador no es la única estrategia disponible para la identificación y la incorporación de los valores y las preferencias de los pacientes en los contenidos de una GPC; existen también los métodos de ***consulta abierta***, ***consulta dirigida***, ***revisiones de literatura*** y *diseños de investigación formal*[21].

Conflictos de interés

El proceso de declaración de intereses para definir la conformación definitiva del grupo desarrollador pretende disminuir el riesgo de que la validez de la guía y la confianza de los usuarios en las recomendaciones formuladas se vean amenazadas por la presencia de conflictos de interés en los integrantes del equipo.

Un conflicto de interés aparece en las circunstancias donde el interés primario de la guía (la validez de los procesos o la seguridad y el bienestar de los pacientes) se ve amenazado por un interés propio del investigador, que puede ser financiero o de prestigio personal o profesional[22]. Con el fin de dar transparencia al proceso y disminuir la influencia de los potenciales conflictos de interés, todos los miembros del grupo desarrollador deben hacer declaraciones de intereses. En la **tabla 22.3** se describen algunas de las actividades que pueden constituir conflictos de interés para el desarrollo de una GPC.

No obstante lo anterior, la identificación y el reporte de un interés relacionado con el tema que abordará la GPC no es suficiente para decidir sobre la participación o la limitación de un potencial miembro del equipo en la formulación de recomendaciones, ni para dar transparencia al proceso de elaboración de la guía[22].

Para el control de potenciales conflictos de interés, es necesario tener: a) el análisis explícito de los intereses declarados, teniendo en cuenta su relación con las tecnologías que serán evaluadas durante la elaboración de la guía, y b) la toma de decisiones sobre la participación o la exclusión de una persona con intereses potencialmente conflictivos en el proceso de formular las recomendaciones de una manera sistemática.

Estos dos procesos promoverán la elaboración de juicios independientes y objetivos por parte del grupo durante los procesos de identificación, selección, evaluación e interpretación de la evidencia, indispensables para formular recomendaciones válidas y confiables[25,26]. Para la conformación del grupo desarrollador y para cualquier proceso de toma de decisiones durante el desarrollo de una GPC, es indispensable llevar a cabo y reportar con claridad y transparencia los procesos de declaración y análisis de intereses. Los documentos que respalden el proceso de identificación de potenciales conflictos de interés, así como las decisiones tomadas con base en ellos, deberán hacer parte del documento final de la guía.

Tabla 22.3 Actividades que pueden constituir conflictos de interés para el desarrollo de una GPC

- Financiación para participar en reuniones académicas, congresos y otros eventos relacionados o no con la profesión (inscripciones, tiquetes aéreos, hospedaje).
- Pago de honorarios por conferencias, ponencias o cursos en reuniones organizadas por la industria de la salud.
- Posesión de acciones o intereses directos en compañías relacionadas con la industria de la salud.
- Financiación para programas de formación.
- Apoyo técnico o financiero para la investigación.
- Empleo en la industria de la salud.
- Prestación actual o previa de servicios a organizaciones gubernamentales relacionadas con el sector salud.
- Trabajo con aseguradores privados o públicos.
- Desarrollo o participación en investigación o tecnologías relacionadas con el tema de interés para la GPC.
- Actividades propias del ejercicio de la profesión de los autores de la GPC.

Fuente: Elaborado a partir de Choudhry[23] y Detsky[24].

Búsqueda de la evidencia

El primer paso en la búsqueda de la evidencia, una vez se tiene claridad sobre las preguntas y el alcance de la guía es la búsqueda de GPC ya elaboradas. Es fundamental buscar la economía de tiempo, de esfuerzos y de recursos. Si se identifican GPC ya elaboradas, se las debe revisar con prioridad, y pueden presentarse varias opciones a ese respecto:

1. **La GPC puede ser adoptada sin cambios:** Para esta decisión, son fundamentales las comparaciones entre las poblaciones donde se obtuvo la evidencia que generó las recomendaciones y la población en la cual se va a proponer la guía; así mismo, es muy importante que las condiciones de disponibilidad y acceso

a recursos sean similares, al igual que las condiciones económicas y logísticas.

2. **La GPC puede ser adaptada:** Puede considerarse que, en principio, la guía puede aceptarse, pero necesita adecuación al medio; básicamente, por diferencias en los puntos ya mencionados: en la población blanco, en la disponibilidad o el acceso a recursos, o en condiciones que la diferencien de las condiciones originales.
3. **La GPC puede ser fuente de información,** aun si no se la puede adoptar o adaptar.

En todos los casos, las **GPC** encontradas deben ser evaluadas críticamente y debe revisarse su actualización. En los dos primeros casos mencionados, debe evaluarse si la **GPC** está actualizada, y si no es así, se la debe actualizar.

Si no hay **GPC** que puedan ser adoptadas o adaptadas, el paso siguiente es la búsqueda de información, un paso fundamental en el desarrollo de **GPC**. Garantizar que esta información sea completa es una pieza fundamental para garantizar la calidad y la confiabilidad de la **GPC**. Por ello, deberán utilizarse métodos explícitos que describan detalladamente los criterios de selección de los estudios, de las bases de datos y de las estrategias de búsqueda empleadas y los años consultados.

Se presenta enseguida una breve descripción de los aspectos que se deben tener en cuenta al hacer la búsqueda de la información. Los lectores interesados podrán ampliar la información al respecto en el primer capítulo del presente libro.

El punto de partida para hacer una búsqueda completa es tener clara la pregunta que se quiere responder, y formularla de tal forma que sea lo suficientemente específica para ser respondida. Se recomienda seguir la estructura ***pacientes-consumidores-participantes, exposición, comparación, resultados*** **(outcomes)**, *tiempo* **(PECOT)**[27], según la cual deben quedar claramente definidas la población, las intervenciones que se van a evaluar, la intervención de referencia con la que se va a comparar, los desenlaces que se tendrán en cuenta para cada una de ellas y el horizonte temporal en el que se evaluará cada uno de ellos **(tabla 22.4)**.

Tener una pregunta formulada de la manera descrita permitirá, adicionalmente, definir los criterios de selección de los estudios clínicos.

Se deben buscar estudios que incluyan a poblaciones similares a la población blanco de la guía, para garantizar que la información será aplicable a las personas en quienes se utilizarán finalmente las recomendaciones (validez externa). Así mismo, deberán quedar claras las intervenciones o las estrategias que se evaluarán, y las cuales pueden ser intervenciones terapéuticas, factores pronósticos, o, incluso, pruebas diagnósticas. Se recomienda para estas últimas considerar el uso de cada prueba como parte de una estrategia diagnóstico-terapéutica[28], buscando evaluar el impacto que tendrá su uso

Tabla 22.4 Componentes de la metodología PECOT

Componentes	Contenido
Pacientes-consumidores-participantes	**¿En cuáles pacientes se está interesado (edad, comorbilidades, estado de** enfermedad)? ¿Se deben contemplar subgrupos de acuerdo con características relevantes?
Exposición	**¿Cuáles intervenciones, tratamientos, factores pronósticos, agentes etiológicos o** pruebas diagnósticas se quiere evaluar? Es necesario incluir todas las alternativas identificadas.
Comparación	**¿Qué alternativa se prevé para comparar con la exposición?: tratamiento habitual o** placebo, ausencia del factor de riesgo o pronóstico, patrón de referencia, etc. De la misma forma, se deben incluir todos los comparadores de interés.
Resultados *(Outcomes)*	¿Qué es en realidad importante para el paciente o consumidor? ¿Qué afecta la exposición? ¿Qué desenlaces deberían tenerse en cuenta?: intermedios o de corto tiempo (baja de la presión arterial), mortalidad, morbilidad, recaídas, readmisiones, regreso al trabajo, actividad física o desempeño social, calidad de vida, estado de salud, desenlaces económicos.
Tiempo	**¿En qué rango de tiempo se espera el desenlace? El rango de tiempo se** recomienda en todos los casos en los que pueda ser relevante.

Fuente: Modificada de New Zealand Guidelines Group 2001[27].

en los desenlaces finales para los pacientes; o sea, en qué medida el uso de las pruebas puede contribuir a la disminución de la mortalidad o de la morbilidad; se entiende que la prueba diagnóstica se encuentra ligada a un determinado plan de tratamiento.

Los desenlaces seleccionados deberán evaluar tanto los beneficios como los riesgos asociados a cada intervención. Deberá darse prioridad a los desenlaces de mayor importancia para los pacientes, teniendo en cuenta desenlaces duros, tales como la mortalidad o la frecuencia de eventos (por ejemplo: infarto, ataque cerebrovascular, fracturas). Otros desenlaces intermedios (medición de niveles sanguíneos de lípidos o de la densidad mineral ósea) podrán ser tenidos en cuenta, pero con menos prioridad que los primeros. Algunos desenlaces adicionales que pueden ser evaluados incluyen los costos o la conveniencia administrativa de una determinada alternativa de manejo.

Con la pregunta estructurada y con los criterios de selección de los estudios definidos, se procederá a realizar la búsqueda de la información. Inicialmente se tendrán en cuenta, como ya se mencionó, otras guías de práctica clínica que hayan sido publicadas sobre el mismo tema. De ellas podrá utilizarse la información o la evidencia en la cual se basan, más que las recomendaciones finales, ya que estas últimas podrán cambiar según consideraciones locales, sociales, culturales o económicas, entre otras[16]. Más adelante en este capítulo se presenta información referente a cómo encontrar GPC.

Posteriormente deberá buscarse la información contenida en los estudios primarios y secundarios (por ejemplo, revisiones sistemáticas de la literatura). La búsqueda de dichos estudios deberá hacerse en diferentes bases de datos. Para la búsqueda de revisiones sistemáticas de la literatura (RSL), se recomienda utilizar el *Index Medicus* (disponible en Internet como MEDLINE), con los filtros definidos para detectar ese tipo de estudio, la base de datos de la Colaboración Cochrane y la herramienta diseñada por el Centro para Revisiones y Diseminación (CRD), diseñada por la Universidad de York (a su vez, disponible en: www.york.ac.uk/inst/crd).

La búsqueda de artículos primarios deberá hacerse, de igual forma, en múltiples bases de datos. Se recomienda incluir bases como Index Medicus y EXCERPTA MÉDICA (disponible en internet como EMBASE), la base de datos de estudios primarios de la Colaboración Cochrane, y bases de datos regionales (LILACS y Redylac, entre otras).

Para llevar a cabo las búsquedas en cualquiera de estas bases de datos, se deberán seleccionar los términos de búsqueda o las palabras clave que representen el tópico. Es conveniente utilizar los términos que se han establecido como MeSH (Medical Subject Heading) y registrar los términos usados en los documentos finales de la guía.

Además de las mencionadas búsquedas, deberán registrarse los intentos adicionales por conseguir información, como la contenida en estudios no publicados o en tesis doctorales, o las aportadas por expertos clínicos.

Evaluación de la calidad de la evidencia

Además de garantizar que la información esté completa, deberá hacerse una evaluación de la calidad de los estudios incluidos, buscando detectar problemas metodológicos que puedan limitar la validez o la aplicabilidad de los resultados. Los lectores interesados podrán ampliar la información en **el capítulo 35** "Apreciación crítica de la literatura".

Para garantizar que esta evaluación sea hecha de una forma completa, sistemática y reproducible, se recomienda utilizar herramientas de evaluación diseñadas para tal fin; si se piensa utilizar información contenida en otras GPC, se recomienda evaluar la calidad de estas con la herramienta AGREE 2[29]; si su calidad resulta aceptable, deberá evaluarse, adicionalmente, la calidad de la RSL que la sustenta. Algunas de las herramientas más utilizadas para evaluar la calidad de RSL son la herramienta AMSTAR[30] y la herramienta PRISMA[31].

Otras escalas son de utilidad para evaluar la calidad de experimentos clínicos aleatorizados: por ejemplo, la presentada en la declaración CONSORT[32], herramientas para evaluar la calidad de estudios de pruebas diagnósticas como el QUADAS[33] o el QUADAS 2[34] o herramientas para evaluar estudios observacionales, como la presentada en la declaración STROBE[35]. Una recopilación actualizada de dichas herramientas está disponible en la página electrónica de la colaboración *Scottish Intercollegiate Guideline Network* (SIGN)[36].

Síntesis de la información

El proceso de síntesis de la información permite identificar similitudes y diferencias entre los estudios incluidos, en cuanto a las poblaciones en las que se desarrollaron y en lo referente a las intervenciones y los desenlaces evaluados.

Cuando la síntesis de la información se hace de forma narrativa, de una forma no estandarizada ni sistemática, el lector no puede establecer si se ha excluido información que podría ser válida y relevante. Por ello, la información suele ser resumida y presentada en tablas de evidencia, que incluyen como sus componentes:

- Descripción del estudio: autor, año de publicación, referencia bibliográfica.
- Características del estudio: diseño, métodos, duración del seguimiento.
- Características de la población.
- Características de la intervención (por ejemplo, dosis de medicamentos, forma de administración o duración del tratamiento en el caso de experimentos clínicos), de la estrategia diagnóstica, o de la forma como se midió el factor de riesgo (en estudios de causalidad).
- Características de la intervención realizada en el grupo de comparación (por ejemplo: placebo, no intervención o ausencia del factor de riesgo).
- Medidas de resultados y síntesis de los resultados principales: incluyendo los desenlaces seleccionados como críticos o importantes para la pregunta, y que deberán incluir tanto los beneficios como los desenlaces de seguridad o los eventos adversos.
- Evaluación de la calidad del estudio: de acuerdo con el análisis hecho utilizando cualquiera de las herramientas diseñadas para tal fin.

Dado que a menudo existen múltiples estudios que abordan un mismo tema, podrá ser necesario hacer una síntesis de los resultados mediante un metaanálisis de la información[37].

Adicionalmente a lo anterior, se ha buscado generar un proceso sistemático que permita sintetizar no solo la evidencia disponible, sino, también, evaluar de forma global la calidad de esta evidencia.

Tradicionalmente, las organizaciones desarrolladoras de guías han utilizado diversas metodologías para evaluar la calidad de la evidencia y la formulación y la graduación de las recomendaciones. En la literatura pueden encontrarse cerca de 50 sistemas diferentes para esta tarea[38]. Sin embargo, se ha reconocido la necesidad de unificar los criterios de evaluación y graduación y la forma de reportar la fuerza de las recomendaciones; particularmente, por la dificultad en la interpretación y por la ineficiencia que representa esta heterogeneidad para los profesionales de la salud[39].

El sistema ***Grading of Recommendations Assessment, Development and Evaluation*** **(GRADE)**[40] ha buscado unificar y estandarizar criterios más completos y explícitos para evaluar la calidad de la evidencia y graduar la fuerza de las recomendaciones. Durante la última década, este abordaje ha sido adoptado y avalado por más de 70 organizaciones internacionales desarrolladoras y compiladoras de **GPC o de evidencia para tomar** decisiones en salud[41]. En Colombia, se cuenta con experiencia en su utilización, ya que es el abordaje recomendado por la ***Guía metodológica*** *para la elaboración de guías de práctica clínica en el Sistema General de Seguridad Social en Salud colombiano*[16] **para la elaboración de guías** y recomendaciones oficiales para el Ministerio de Salud y Protección Social.

El abordaje **GRADE se diferencia de otros sistemas** para graduar la evidencia porque para su evaluación no considera de forma individual los estudios, sino que reúne en una sola calificación toda la información disponible (proveniente de uno o de múltiples estudios) para cada uno de los desenlaces críticos para la toma de las decisiones en la formulación de recomendaciones.

De acuerdo con lo propuesto por dicho sistema, existen cinco factores que podrían disminuir la calidad de la evidencia partiendo del diseño de los estudios individuales, y tres factores que podrían aumentarla **(tabla 22.5)**. **La calificación** final para cada desenlace se encontrará en alguna de las cuatro categorías definidas: ***alta***, ***moderada***, *baja* o ***muy baja***[40]. **El significado de cada una de** estas categorías se describe en la **tabla 22.6**.

Generación de recomendaciones

Una vez el grupo desarrollador ha sintetizado la evidencia (o ha conformado el cuerpo de la evidencia que servirá de base para una

Tabla 22.5 Criterios de evaluación de calidad del conjunto de la evidencia

Factores que disminuyen la calidad	Descripción
Riesgo de sesgo	Los estudios (ECA y observacionales) pueden arriesgarse a presentar resultados sesgados si tienen fallas en el diseño o la conducción[42]. Elementos para evaluar en ECA con el fin de determinar el riesgo de sesgo: • Enmascaramientos. • Manejo de datos faltantes. • Sesgo de reporte selectivo de desenlaces. • Otros. Elementos por evaluar en estudios observacionales para determinar el riesgo de sesgo: • Criterios de elegibilidad de la población. • Métodos de medición de la exposición y del desenlace. • Control de confusión. • Seguimiento incompleto.
Inconsistencia	**Debe considerarse que los resultados son inconsistentes si[43]:** • Los estimativos puntuales varían ampliamente entre los estudios. • Los índices de confianza (IC) **no se sobreponen.** • Las pruebas estadísticas de heterogeneidad muestran un valor de *p* bajo. • El valor I^2 **es alto.**
Evidencia indirecta	La evidencia podría ser indirecta si la intervención, la población o los desenlaces evaluados son diferentes de los de interés en la GPC. GRADE considera cuatro razones principales de evidencia indirecta[44]: • Diferencias en la población (aplicabilidad). • Diferencias en la intervención (aplicabilidad). • Diferencias en los desenlaces (sustitutos). • Comparaciones indirectas.
Imprecisión	La amplitud del intervalo de confianza en torno al estimativo del efecto podría llevar a disminuir la calidad de la evidencia en uno o dos niveles (dependiendo de su extensión)[45].
Sesgo de publicación	La síntesis de los estudios puede proveer estimativos sesgados si los autores de la RSL no detectan la totalidad de los estudios que abordan la pregunta de interés[46].
Factores que aumentan la calidad	**Descripción**
• Gran magnitud del efecto. • Presencia de gradiente dosis/respuesta. • Los factores de confusión presentes disminuirían el efecto encontrado.	Los factores que podrían aumentar la calidad de la evidencia son poco comunes, y se consideran de relevancia, principalmente, para estudios observacionales. No se recomienda utilizar dichos criterios cuando haya disminuido la evaluación de calidad por riesgo de sesgo[47].

Fuente: Adaptado de Guyatt[40,42-46].

Tabla 22.6 Significado de los cuatro niveles de evidencia dentro del abordaje GRADE

Calidad de la evidencia	Definición	Representación gráfica
Alta	**Se tiene gran confianza en que el verdadero efecto se encuentra** cerca del estimativo del efecto.	⊕⊕⊕⊕
Moderada	**La confianza en el estimativo del efecto es moderada: es probable** que el verdadero efecto se encuentre cerca del estimativo, pero existe la posibilidad de que sea sustancialmente distinto.	⊕⊕⊕
Baja	**La confianza en el estimativo del efecto es limitada: el verdadero** efecto puede ser sustancialmente distinto del estimativo del efecto.	⊕⊕
Muy baja	**La confianza en el estimativo del efecto es muy baja: es probable** que el verdadero efecto sea sustancialmente distinto del estimativo del efecto.	⊕

Fuente: Traducido Balshem H. et al.[48].

recomendación), y una vez ha evaluado su calidad, podrá proceder con el proceso de formular recomendaciones.

Las recomendaciones son las afirmaciones que apoyan a los profesionales de la salud y a los pacientes en la toma de decisiones respecto a las alternativas más seguras y efectivas para la prevención, el diagnóstico, el tratamiento o la rehabilitación en caso de una condición o de una enfermedad en específico.

Una de las principales novedades presentadas por el abordaje GRADE respecto a los sistemas utilizados tradicionalmente es la separación de la calidad de la evidencia y la fuerza de la recomendación como elementos independientes, de tal forma que la disponibilidad de evidencia de alta calidad no necesariamente implica formular recomendaciones fuertes[39]. Además de la calidad de la evidencia, el sistema GRADE incorpora la consideración de elementos adicionales para graduar la fuerza de una recomendación[49]:

- El balance entre los beneficios y los riesgos o cargas esperados y la administración de una alternativa particular.
- La variabilidad o la confianza que se tengan en los valores y las preferencias de los pacientes respecto a dicha alternativa.
- El uso de recursos, asociado a la implementación de una recomendación particular.

Tomando en cuenta dichos elementos, las recomendaciones pueden formularse a favor o en contra de una alternativa particular, con dos grados diferentes de seguridad: ***fuerte*** o ***débil*** (condicional) **(tabla 22.7)**.

Los procesos para formular recomendaciones se desarrollan en reuniones donde participa la totalidad del grupo desarrollador, y, generalmente, se basan en metodologías de consenso formal o informal, de manera variable. En dichas reuniones se pretende que el grupo llegue a acuerdos sobre cuál evidencia será utilizada para basar la formulación de una recomendación particular, su calidad, y sobre las consideraciones adicionales que determinarán la dirección y la fuerza de una recomendación.

Los grupos desarrolladores deben reportar los procesos que fueron llevados a cabo para la formulación de recomendaciones, describir la metodología utilizada y los acuerdos o los desacuerdos que se hayan presentado en el grupo respecto a la toma de la decisión para formular cada recomendación. La presentación transparente de los procesos y de los consensos que se hayan hecho, y las discusiones generadas en torno a cada una de las variables que determinan la dirección y la fuerza de una recomendación aportan claridad y transparencia a un proceso que, por su naturaleza, se halla sujeto a juicios por parte de los integrantes del equipo desarrollador.

Implementación de las GPC

Para que una guía alcance el impacto deseado, se debe lograr una exitosa implementación de las recomendaciones en la práctica clínica.

Tabla 22.7 Fuerza, representación gráfica y significado de una recomendación según el abordaje GRADE

Fuerza y dirección de una recomendación	Representación gráfica	Significado
Recomendación fuerte a favor	↑↑	**Una recomendación fuerte implica que la** mayoría de los pacientes adecuadamente informados escogerá la alternativa o el curso de acción recomendado, y que los profesionales de la salud recomendarán la alternativa.
Recomendación fuerte en contra	↓↓	
Recomendación débil a favor	↑?	**Una recomendación débil implica que la** elección de los pacientes será variable y dependerá de sus preferencias y los valores personales. Los profesionales de la salud tomarán la decisión sobre la recomendación o no de una alternativa dependiendo de las preferencias y las circunstancias particulares de cada paciente.
Recomendación débil en contra	↓ ?	

Fuente: Elaborado a partir de Guyatt[50] y Schünemann[51].

La implementación de las recomendaciones de una guía puede involucrar acciones de gran alcance en diferentes planos dentro un sistema de salud: toma de decisiones sobre su implementación en todo un país, entidades gubernamentales regionales y locales, IPS, empresas promotoras y profesionales de la salud. En algunas ocasiones, las GPC se desarrollan o se implementan en escenarios limitados: únicamente en instituciones de salud, privadas o públicas, con el fin de estandarizar los procesos de atención y alcanzar mejores niveles en la calidad en la atención. Cualquiera que sea el escenario, las actividades relacionadas con la difusión de la información y con la apropiación y el uso de las recomendaciones por parte de los profesionales de la salud implica diseñar y ejecutar un plan de implementación.

Generalmente, los grupos desarrolladores no están a cargo de crear ni de ejecutar los planes de implementación para las recomendaciones; las actividades requeridas y relacionadas con dicho proceso suelen estar fuera del alcance de las decisiones y del campo del acción del equipo. Sin embargo, existen características intrínsecas de las guías que podrían facilitar su implementación[27,52]. Los grupos deben desarrollar la guía, formular las recomendaciones, y elaborar los documentos tomando en cuenta las características que aumentan la probabilidad de que una GPC sea implementada con éxito (tabla 22.8).

Cómo encontrar GPC



En la actualidad, tenemos disponibles múltiples herramientas para identificar y tener acceso a GPC.

Tabla 22.8 Características de una GPC que favorecen la implementación exitosa de las recomendaciones

Característica	Descripción
Validez	**Se debe describir y reportar con claridad el listado de referencias en el que se basa** cada recomendación, y utilizar un sistema consistente de evaluación de calidad de evidencia y graduación de recomendaciones.
Aplicabilidad	Las recomendaciones deben incluir toda la información relevante para facilitar su comprensión y su aplicación a pacientes particulares. Deben proveerse cuadros o textos explicativos de las recomendaciones cuando se requiera, o cuando haya condiciones especiales para su uso.
Versatilidad	**Se deben proveer diferentes versiones del documento: para profesionales de la salud,** versiones de consulta rápida, aplicativos móviles, documentos para pacientes; también, aumentar su difusión publicando el contenido en revistas científicas.
Facilidad en la utilización	**Los documentos deben ser fácilmente navegables, con tablas de contenido claras y** completas. Se deben incluir figuras o algoritmos que expliquen el flujo de atención o la relación de las recomendaciones entre sí.
Alcance y escenario	Debe ser claro en el documento de la guía cuál es el alcance de las recomendaciones y a qué tipo de profesionales de la salud se encuentran dirigidas. Puede ser de gran ayuda incluir los requisitos estructurales para implementar las recomendaciones, los requerimientos de capacitación o la estandarización de procedimientos. Las implicaciones sobre los recursos deben ser discutidas y descritas en el documento.
Contacto y comunicación	La guía debe proveer recursos para encontrar información adicional. Para los profesionales de la salud es importante incluir datos de contacto del equipo desarrollador y acceso a la documentación de soporte; para los pacientes o los cuidadores, proveer fuentes adicionales de información, puntos de atención o recursos educativos.
Elementos para la implementación	Los grupos pueden describir un ejercicio de identificación de barreras y facilitadores para la implementación: presentaciones, listados de recomendaciones y figuras, con el fin de que se hagan actividades de difusión y educación en torno a las recomendaciones en el ámbito local.
Evaluación	**Se refiere a la inclusión de indicadores propuestos como claves para monitorizar la** implementación de las recomendaciones y el impacto de su uso.

Fuente: Elaborado a partir de NZGG[27] y Gagliardi[52].

Se destacan los instrumentos desarrollados por el Departamento de Salud y Servicios Humanos de los Estados Unidos de América (NGC, por las siglas en inglés del National Guideline Clearinghouse) y por la colaboración Guideline International Network (GIN).

Además de lo anterior, las GPC se pueden buscar directamente en las páginas electrónicas de grupos internacionales reconocidos por producir documentos de esta clase de alta calidad. Dentro de ellos se destacan las siguientes fuentes: New Zeland Guidelines Group (NZGG), National Institute for Clinical Excellence (NICE) y Scottish Intercollegiate Network (SIGN).

Por último, cabe resaltar fuentes que permiten identificar guías en lengua española. Se destacan: Guía Salud (España), Redsalud (Chile) y Cenetec (México). La información sobre la forma de acceder a las diferentes fuentes de información se presenta en la **tabla 22.9**.

Tabla 22.9 Fuentes de GPC

Tipo de fuente	Fuente	Dirección electrónica
Organismos recopiladores de guías	National Guideline Clearinghouse (NGC)	**http://www.guideline.gov/**
	Guideline International Network (GIN)	**http://www.g-i-n.net/**
	Catalogage et l'Indexation des Sites Médicaux (CISMEF).	**http://www.chu-rouen.fr/cismef/**
	German Agency for Quality in Medicine	www.aezq.de/
Organismos que producen GPC	New Zeland Guidelines Group (NZGG)	**http://www.nzgg.org.nz**
	National Institute for Clinical Excellence (NICE)	**www.nice.org.uk/Guidance/Topic**
	Scottish Intercollegiate Network (SIGN)	**http://www.sign.ac.uk**
GPC nacionales o iberoamericanas	Guía Salud	http://www.guiasalud.es/home.asp
	Redsalud	http://www.redsalud.gov.cl/gesauge/ges_descargas.html
	Cenetec	**www.cenetec.salud.gob.mx/interior/** catalogoMaestroGPC.html
Bases de datos generales	PubMed	http://www.ncbi.nih.gov/entrez/query.fcgi
	Medline	http://www.ncbi.nih.gov
	EMBASE	**http://www.embase.com/**
	TripDatabase	http://www.tripdatabase.com/index.html

Fuente: Elaboración propia.

Referencias

1. Field M, Lohr K. Clinical practice guidelines: Directions for a new agency. Medicine I of, editor. Washington DC: National Academic Press; 1990.
2. Gómez-Restrepo C. Redalyc. Guías de práctica clínica: ¿cuáles son sus alcances? Rev Colomb Psiquiat. 2010;39:8-10.
3. Bonfill X, Marzo M. Guías de práctica clínica: tenerlas, que sean de calidad y que salgan del armario. Med Clínica. 2003;120:496-7.
4. Bonfill X. Algunos apuntes para impulsar una atención sanitaria basada en la evidencia. Med Clin Barc. 2002;118(Supl 3):2-6.
5. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. J Assoc Medicale Can. 1997;157:408-16.
6. Upshur REG. Do clinical guidelines still make sense? No. Ann Fam Med. 2014;12:202-3.
7. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet. 1993;342:1317-22.
8. Sox HC. Do clinical guidelines still make sense? Yes. Ann Fam Med. 2014;12:200-1.
9. Tunis SR, Hayward RS, Wilson MC, et al. Inter-

nists' attitudes about clinical practice guidelines. Ann Intern Med. 1994;120:956-63.
10. Woolf SH. Practice guidelines, a new reality in medicine. Arch Intern Med. 1992;152:946.
11. Medicine I of. Clinical practice guidelines we can trust. Press NA, editor. Washington DC; 2011.
12. Moynihan R, Oxman AD, Lavis JN PE. Evidence-informed health policy: Using research to make health systems healthier [internet] 2006 [citado 2014 jul. 18]. Disponible en: http://www.researchgate.net/publication/27514681_Evidence-Informed_Health_Policy_using_research_to_make_health_systems_healthier
13. Viergever RF, Olifson S, Ghaffar A, et al. A checklist for health research priority setting: nine common themes of good practice. Health Res Policy Syst. 2010;8:36.
14. Oxman AD, Schünemann HJ, Fretheim A. improving the use of research evidence in guideline development: 2. Priority setting. Health Res Policy Syst. 2006;4:14.
15. Reveiz L, Téllez DR, Castillo JS, et al. Prioritization strategies in clinical practice guidelines development: a pilot study. Health Res Policy Syst. 2010;8:7.
16. Carrasquilla G, Pulido A, de la Hoz AM, et al. Guía Metodológica para la elaboración de Guías de Práctica Clínica con Evaluación Económica en el Sistema General de Seguridad Social en Salud Colombiano. Bogotá: Minsalud; 2014.
17. SIGN. Sign 50 A guideline developer's handbook [Internet]. Edinburgh: Scottish Intercollegiate Guidelines Network; 2011. Available from: www.sign.ac.uk
18. Chong C, Chen I, Naglie G, et al. How well do guidelines incorporate evidence on patient preferences? J Gen Intern Med. 2009;24:977-82.
19. Légaré F, Boivin A, van der Weijden T, et al. Patient and public involvement in clinical practice guidelines: a knowledge synthesis of existing programs. Med Decis Making. 2011;31:E45-74.
20. Van Wersch A, Eccles M. Involvement of consumers in the development of evidence based clinical guidelines: practical experiences from the North of England evidence based guideline development programme. Qual Health Care. 2001;10:10-6.
21. G-I-N PUBLIC. G-I-N PUBLIC Toolkit : Patient and Public. Guidelines International Network. Berlin: GIN Public; 2012.
22. Thompson D. Understanding financial conflicts of interest. N Engl J Med. 1993;329:573-6.
23. Choudhry NK DAS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA. 2002;287:612-7.
24. Detsky A. Sources of bias for authors of clinical practice guidelines. Can Med Assoc J. 2006;179:1033.
25. NICE. The guidelines manual [Internet]. Londres: National Institute for Health and Clinical Excellence; 2012.
26. Hurst SA, Mauron A. The ethics of managing conflicts of interest. EMBO Rep. 2008;9:119.
27. NZGG. Handbook for the preparation of explicit evidence-based clinical practice guidelines. Wellington: New Zealand Guidelines Group; 2001.
28. Schumemann A, Oxman A, Brozek J. GRADE : grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008;336.1106.
29. AGREE. Appraisal of Guidelines for Research & Evaluation II Instrument. The AGREE Next Steps Consortium. Hamilton, Ontario, Canada; 2009. p. 1-57.
30. Shea BJ, Grimshaw JM, Wells G, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
31. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
32. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010;63:e1-37.
33. Whiting P, Rutjes AWS, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25.
34. Whiting P, Rutjes AWS, Westwood M. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
35. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. Int J Surg. 2014;335:20-2.
36. Network SIG. Sign 50. Edinburgh: Scottish Intercollegiate Guidelines Network; 2011.
37. TCCI. Manual Cochrane de revisiones sistemáticas de intervenciones, versión 5.1.0 [actualizada en marzo de 2011]. 2011.
38. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evid Report/Technology Assess Vol AHRQ Publ. 2002;02-E016.

39. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924-6.
40. Guyatt G, Oxman AD, Akl E, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64:383-94.
41. Gradeworkinggroup.org [internet]; The GRADE working group; c2005-2014 2002 [citado 2014 oct. 28]. Disponible en: http://www.gradeworkinggroup.org/society/index.htm
42. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21247734
43. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol. 2011;64:1294-302.
44. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011;64:1303-10.
45. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64:1283-93.
46. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol. 2011;64:1277-82.
47. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011;64:1311-6.
48. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401-6.
49. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66:726-35.
50. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. BMJ. 2008;336:1049-51.
51. Schünemann H, Brozek J, Oxman AE. GRADE handbook for grading quality of evidence and strength of recommendation. Version 3.2. Version 32 [updated March 2009] GRADE Work Group. The GRADE Working Group, 2009. Disponible en: http://www.cc-ims.net/gradepro.
52. Gagliardi AR, Brouwers MC, Palda V, et al. How can we improve guideline use? A conceptual framework of implementability. Implement Sci. 2011;6:26.



SECCIÓN 5

Análisis económicos y evaluación de la tecnología

CAPÍTULO 23 Introducción a la evaluación económica de tecnologías en salud
Hoover Quitián Reyes

CAPÍTULO 24 Estudios de farmacoepidemiología y farmacoeconomía
Rodolfo José Dennis Verano

CAPÍTULO 25 Evaluaciones de tecnología
María Ximena Rojas Reyes

CAPÍTULO

23

Introducción a la evaluación económica de tecnologías en salud

HOOVER QUITIÁN REYES

En la práctica clínica se tiene acceso a múltiples intervenciones, farmacológicas y no farmacológicas, con efectividades diversas y casi siempre buena seguridad, que son bien conocidas por el clínico o por el tomador de decisiones. Pero con frecuencia se desconocen los aspectos económicos de dichas intervenciones, o no se hace el ejercicio de poner en la balanza los riesgos y costos, sopesados con los beneficios potenciales. Como una forma de demostrar la utilidad para la toma de decisiones en clínica, un ejemplo clásico evaluó la prevención de riesgo cardiovascular con una intervención para bajar el colesterol.

EJEMPLO

Cada año de vida salvado en hombres y mujeres de muy alto riesgo cardiovascular, con edades entre los 45 y los 64 años, y que reciben una estatina como tratamiento para bajar el colesterol LDL, vale £6000, tal como se determinó en un estudio con modelamiento[1]. En población de riesgo alto vale £32 000; en población de riesgo bajo, £136 000, y en población de riesgo muy bajo, £361 000. Es claro en el estudio que la misma medicación en el mismo número de personas puede producir resultados muy diferentes según el nivel de riesgo. Esta información ayuda en la toma de decisiones, tanto a médicos como a decisores, en políticas en salud.

¿POR QUÉ HACER EVALUACIONES ECONÓMICAS?

La economía es la ciencia social que estudia las necesidades del ser humano y cómo satisfacerlas dados los escasos recursos existentes[2].

En este contexto, la ***escasez*** se refiere a la imposibilidad de emplear un determinado recurso para suplir necesidades de manera adecuada e ininterrumpida, sin que este, eventualmente, se agote. Se debe entender por ***recursos*** no solo la capacidad de financiamiento, sino también, el capital humano, la tecnología, la infraestructura y los elementos de la naturaleza necesarios para la producción de bienes o para la prestación de servicios.

La escasez es un concepto dinámico que depende de dos magnitudes, las cuales pueden evolucionar de manera independiente: 1) la cantidad disponible del recurso (***oferta***) y 2) la cantidad requerida de dicho recurso (***demanda***) para satisfacer cierta necesidad.

EJEMPLO

El aire es un recurso vital, y antes de la segunda mitad del siglo XX no se hacían análisis económicos relacionados con él, debido a que se lo consideraba ilimitado. Sin embargo, con el crecimiento de la contaminación y la consecuente disminución del aire apto para el adecuado funcionamiento del cuerpo humano, en la actualidad, incluso, se han generado mercados donde los responsables de la contaminación hacen un pago para compensar el daño causado, o donde quienes hacen uso de algunas instalaciones deben incurrir en costos adicionales para contar con purificadores del ambiente[3,4].

Aunque la prestación de servicios en salud tiene una mayor incertidumbre e involucra elementos éticos que no existen en otros sectores[5], sí comparte muchos elementos con las demás actividades productivas. Atender la salud de personas y comunidades es un proceso productivo que consume recursos: insumos, personal médico, personal de enfermería, personal administrativo, medicamentos, elementos tecnológicos e infraestructura hospitalaria, entre otros, que interactúan dentro de un sistema para generar productos y servicios (atenciones en salud) en las áreas de clasificación (tamización, búsqueda de casos, diagnóstico, pronóstico) e intervención: promoción de la salud, prevención primaria, secundaria (tratamiento) terciaria y rehabilitación. Estos productos y servicios asistenciales buscan generar bienes: prevenir la muerte prematura, y alcanzar o mantener un estado adecuado de salud, con el mantenimiento o el mejoramiento consecuentes de la calidad de vida relacionada con salud.

Específicamente en el sector salud, pese a la tendencia mundial a incrementar los recursos destinados a atender problemas sanitarios, el crecimiento de la demanda y la necesidad de emplear cada vez más recursos no solo han impedido que la escasez disminuya, sino que en numerosos casos la han incrementado[6-8]

Hay dos elementos que explican el alza en la demanda de servicios en salud, y que, paradójicamente, son consecuencia de los avances en la tecnología sanitaria. El primero de ellos es el éxito de las intervenciones: la mejora en la tecnología sanitaria ha generado un incremento de la esperanza de vida, y al vivir más años aumenta la probabilidad de tener que acudir al médico por enfermedades crónicas, por problemas de salud propios del proceso de envejecimiento o por exposición a factores medioambientales nocivos.

El segundo elemento es la mayor complejidad de la práctica médica. Conforme se avanza en el conocimiento de las enfermedades, aparecen más y mejores alternativas para su detección o su tratamiento. Ello implica presión sobre los recursos, no solo porque existen más tecnologías, sino también, porque estas nuevas intervenciones requieren a profesionales más especializados, cuyos honorarios son mayores, y debe contarse con la infraestructura necesaria para su formación y para su ejercicio profesional. Adicionalmente, la dinámica misma de la innovación abre nuevas líneas de investigación que involucran cada vez más recursos, por lo cual los poseedores de los derechos de propiedad quieren recuperar el valor de su inversión, y eso, a su vez, lleva a que la nueva tecnología, en la mayoría de los casos, se ofrezca a precios mayores, aun si resulta menos costosa que las anteriores.

En síntesis, el problema central de la economía de la salud puede formularse en términos de la siguiente pregunta: ***¿Cómo se pueden asignar*** *de forma óptima los recursos disponibles (escasos, costosos y finitos) para producir bienes y servicios en cantidad y calidad suficientes para satisfacer las aparentemente ilimitadas necesidades de atención en salud de la sociedad?*

Los criterios para asignar recursos escasos y finitos en una sociedad equitativa incluyen aspectos sociales (ética), técnicos (eficacia, efectividad, calidad), políticos (democracia, equidad), personales (utilidad: preferencias y valores individuales) ***y económicos (eficiencia)***

La valoración económica del proceso de asignación de recursos no reemplaza ni supera ninguno de los otros criterios mencionados, pero sí lo suplementa y lo racionaliza.

La economía de la salud propone hacer un proceso explícito y sistemático de toma de decisiones al momento de distribuir los recursos finitos (escasos) para atender las aparentemente ilimitadas necesidades de atención en salud, que permita asignar racionalmente los recursos para minimizar el llamado ***costo de oportunidad***, concepto que se explica seguidamente.



En un escenario donde los recursos son limitados o finitos, deben tenerse en cuenta consideraciones de orden económico por la posibilidad de escasez[9,10] del capital humano, físico o financiero necesario para la prestación de servicios en salud.

Si existiera un escenario hipotético en el cual los recursos fuesen ilimitados, las decisiones solo tendrían en cuenta la efectividad y la seguridad de las intervenciones.

Dado que los recursos no son ilimitados, se hace necesario considerar el costo de oportunidad de cada una de las decisiones. Al ser insuficientes los recursos, efectuar cierta intervención puede significar la no realización de otras actividades que podrían mejorar la salud del mismo paciente o la de otros integrantes de la sociedad; resulta indispensable evaluar si la magnitud de los beneficios en salud derivados de una determinada intervención justifican incurrir en el gasto necesario para llevarla a cabo, o si resultaría mejor darle otros usos al recurso. Si en una situación específica las alternativas A y B compiten por los recursos para realizarlas, y se opta por la alternativa A, el costo de oportunidad corresponde al valor del bien o el servicio que habría generado la alternativa B, en caso de haberse destinado los recursos a ella.

EJEMPLO

Un tratamiento complejo y por el que se debe pagar una elevada suma de dinero puede prolongar tan solo algunos días la vida de un paciente con cáncer avanzado; como máximo, tres semanas. A primera vista, es evidente que se extiende la vida de una persona (idealmente, en condiciones favorables), por lo cual se justificaría aplicar el tratamiento; sin embargo, la decisión no es tan clara si se evalúa el costo de oportunidad: el número que vidas que podrían salvarse si se emplease dicho dinero para otros propósitos.

De igual forma, pensando tan solo en su paciente, el médico, tal vez, decida hacer una consulta médica con interrogatorio y examen físico únicamente después de contar con una batería completa de exámenes. Sin embargo, si tiene en cuenta en el proceso de decisión los recursos que se necesitarían para hacer los mismos análisis a todos los pacientes de la misma edad antes de la consulta, es posible que se decida establecer como protocolo de trabajo hacer primero la consulta médica, y con la información obtenida, ordenar exámenes exclusivamente a quienes tengan indicación por hallazgos en la consulta o por tener factores de riesgo. Así mismo, ordenaría solo los exámenes que estén justificados.

¿QUÉ SON LAS EVALUACIONES ECONÓMICAS?

Las ***evaluaciones económicas*** son análisis que buscan establecer el valor monetario de los recursos empleados en la utilización de tecnologías sanitarias y su correspondencia con los resultados en salud obtenidos.

Dependiendo del objetivo de la evaluación, la metodología para aproximarse a los valores monetarios, así como la empleada para estimar los resultados obtenidos en salud, puede variar[11].

La efectividad y los costos del uso de una tecnología dependen del contexto. Por tal motivo, es necesario diferenciar las tecnologías en salud no solo por sus características físicas, químicas y operativas, sino también, por tiempo y lugar. Si bien la técnica para un urocultivo se encuentra estandarizada, así como los insumos requeridos, será de fácil acceso en una ciudad capital, pero puede requerir una logística diferente en regiones apartadas (traslado a otra ciudad, condiciones de transporte, refrigeración en algunos casos…); este tipo de restricciones, sin embargo, podrían desaparecer con el tiempo.

De igual forma, la efectividad de las intervenciones puede verse afectada por las características epidemiológicas, el biotipo o, incluso, la cultura de la población.

EJEMPLO

La suplementación a menores de 5 años con vitamina A o con zinc puede ofrecer grandes beneficios en regiones donde la dieta no incluya alimentos que aporten los niveles mínimos requeridos. Esta misma suplementación será claramente innecesaria, aunque sea muy económica, en regiones o en poblaciones donde la alimentación sea apropiada y suficiente.

Además de lo anterior, la forma como se organice la sociedad para cubrir los servicios de salud, ya sea mediante sistemas de aseguramiento (parciales, totales, personales o colectivos) determinará, en gran medida, los costos que deban incluirse.
En consecuencia, las evaluaciones económicas, por definición, no son generalizables.

Los resultados obtenidos en otras regiones o naciones no pueden asumirse como válidos en otras latitudes, y tampoco es posible considerar que las conclusiones serán útiles por mucho tiempo[10,12,13].

Debido a la complejidad del análisis, hacer evaluaciones económicas requiere un equipo multidisciplinario que involucre a expertos clínicos, para asegurar que los supuestos que se proponen mantienen la verosimilitud del problema médico, y a profesionales de disciplinas afines a las ciencias económicas que comprendan el funcionamiento del sistema de salud donde se lleva a cabo el estudio.

¿Cuándo es posible hacer evaluación económica a tecnologías en salud?

Cada uno de los tipos de estudios económicos tiene diferentes condiciones necesarias para su adecuado desarrollo. En general, la evaluación económica de tecnologías en salud requiere un conjunto de información de calidad suficiente, pese a que es posible emplear supuestos o acudir al consenso de expertos para construir los valores de los parámetros requeridos. Cuanto mayor sea el número de variables definidas con información que no provenga de literatura científica de buena calidad, menor será la validez de los resultados obtenidos.

¿QUÉ TIPO DE EVALUACIONES ECONÓMICAS EXISTEN?

Existen varios tipos de análisis económicos que apoyan la toma de decisiones sobre tecnología en salud, y cada uno de ellos responde a interrogantes distintos empleando diversas metodologías. En principio, se pueden dividir en evaluaciones ***parciales*** y ***completas*** **(tabla 23.1).**

Las evaluaciones de tecnología pueden ser parciales o completas dependiendo de si se comparan dos o más alternativas, y de si se contrastan tanto los costos de los recursos y el proceso productivo como los beneficios derivados de los bienes, los productos y los servicios generados. Una evaluación completa de una intervención en salud debe identificar, cuantificar, valorar y comparar los costos y los beneficios de las alternativas de interés. Cuando se comparan tanto los costos como los beneficios en dos o más alternativas, se trata de una evaluación completa. Si falta cualquiera de los elementos en al menos una alternativa, o si solo se examina una alternativa, se trata de una evaluación parcial[10]. **Basar las decisiones** sobre asignaciones de recursos únicamente en evaluaciones parciales es arriesgado y no recomendable.

En evaluación de tecnologías sanitarias se emplean, de forma complementaria y depen-

Tabla 23.1 Características que diferencian las evaluaciones económicas parciales y completas

		¿Se examinan recursos (costos) y desenlaces?		
		No		Sí
		Solo desenlaces	**Solo** costos	
¿Se compararon dos o más alternativas?	No	**Evaluación parcial**		**Evaluación parcial**
		Describe desenlaces	Describe costos	Describe costos y desenlaces
	Sí	Evaluación parcial		**Evaluación completa** • Costo-minimización • Costo-efectividad • Costo-utilidad • Costo-beneficio
		Estima efectos (eficacia efectividad)	**Analiza costos**	

Fuente: Modificado de Drummond M., et al.[10].

diendo de los objetivos específicos, al menos seis tipos de estudios económicos **(tabla 23.2)**. Los dos primeros solo miran aspectos del costo (costo de enfermedad) o del gasto (impacto presupuestal), y corresponden a evaluaciones parciales; los cuatro últimos contrastan los valores de recursos y las consecuencias entre varias alternativas, y corresponden a evaluaciones completas.

Evaluaciones parciales

Estudio del costo de enfermedad

EJEMPLO

Un estudio en 2004[14] mostró que los costos agudos de los ataques cerebrovasculares (ACV) variaban entre US$530 y US$32 370 al evaluar a 140 pacientes con ACV isquémicos y hemorrágicos. Se encontró que había interrelaciones importantes entre el tipo de ACV y la etiología con la longitud de hospitalización, con los costos y con la fatalidad a 30 días. La mayoría de los predictores univariados de uso de recursos, costos y desenlaces tuvieron efecto de confusión por la gravedad inicial del ACV.

Un estudio de costo-enfermedad permite establecer el comportamiento económico de la enfermedad; además, determina cuánto cuesta convivir con ella, tratarla y generar programas de prevención. Es necesario construir la historia natural de la enfermedad e identificar todos los costos asociados; no solo los relacionados con el tratamiento: también deben involucrarse la prevención, el diagnóstico y los costos del manejo de los posibles desenlaces. En los estudios de costo de enfermedad también puede tratar de atribuirse un valor monetario a las consecuencias del proceso patológico, lo que puede convertir a dicho estudio en la contraparte no comparativa (parcial) de un estudio de costo-beneficio.

Tabla 23.2 Tipos de evaluaciones económicas

- Estudio del costo de enfermedad
- Estudio de impacto presupuestal
- Minimización de costos
- Estudio de costo-efectividad
- Estudio de costo-utilidad
- Estudio de costo-beneficio

Impacto presupuestal

El impacto presupuestal es analizado una vez haya evidencia de eficacia y seguridad de una tecnología.

Busca responder cuánto costaría su cubrimiento dentro del sistema de salud, cualquiera que este sea.

En primera instancia, se debe calcular el número de personas que recibirían la intervención, y dicho valor se multiplica por el costo esperado total de proporcionar la tecnología sanitaria, según las condiciones propias del sistema de aseguramiento. Por ejemplo, posiblemente no se requiera la inclusión del costo de ciertos insumos, por cuanto estos, de acuerdo con la normatividad, deben ser asumidos por el paciente. El costo esperado total incluye las tecnologías complementarias requeridas para el correcto uso de la intervención evaluada, así como los servicios generados por el tratamiento de los posibles efectos adversos durante el horizonte temporal adecuado, aunque usualmente los resultados se presentan anualizados, para que sean más útiles, pues las finanzas públicas a menudo son planeadas para un periodo de un año. Adicionalmente, si hay más alternativas, se construyen varios escenarios plausibles sobre la posible distribución de la demanda de servicios, debido a que no necesariamente todos los pacientes serán tratados o diagnosticados con la misma alternativa[14]. Aun si hay más de una alternativa, el objetivo primario de ese estudio no es determinar la eficiencia del programa o su implementación, sino proyectar el gasto requerido, con la perspectiva de un financiador específico (por ejemplo, el Estado).

Evaluaciones completas

Las evaluaciones completas pueden llevarse a cabo simultáneamente y como parte de un estudio experimental de efectividad y seguridad, o pueden conducirse sobre revisiones sistemáticas y modelos que estimen la efectividad y la seguridad a partir de fuentes primarias válidas y confiables. En todos los casos se contrastan los costos y las consecuencias de alternativas enfrentadas **(tabla 23.3)**.

Tabla 23.3 Supuestos e indicadores de resultados de análisis económicos que contrastan costos y consecuencias de intervenciones en salud

Tipo de análisis	Supuestos sobre costos	Supuestos sobre desenlaces (efectos)	Indicador del estudio
Minimización de costos	Costos no establecidos; pueden ser diferentes entre intervenciones. Se enumeran los recursos y se les asignan valores monetarios estandarizados (por ejemplo, en pesos colombianos de 2014).	Efectos ya establecidos; deben ser equivalentes (efectividad y seguridad, habilidad discriminativa, etc.)	Diferencia de costos
Costo-efectividad	Costos no establecidos; pueden ser diferentes entre intervenciones. Se enumeran los recursos y se les asignan valores monetarios estandarizados (por ejemplo, en pesos colombianos de 2014).	Se espera que los efectos sean diferentes. La efectividad se mide utilizando una sola unidad. Puede ser el efecto sobre un desenlace (infarto prevenido) o sobre un índice combinado de morbilidad y mortalidad (por ejemplo, años de vida saludable ajustados por discapacidad).	Razón incremental de costo-efectividad
Costo-Utilidad	Costos no establecidos; pueden ser diferentes entre intervenciones. Se enumeran los recursos y se les asignan valores monetarios estandarizados (por ejemplo, en pesos colombianos de 2014).	Se espera que los efectos sean diferentes. Los efectos positivos y negativos sobre el padecimiento y sobre la salud del paciente se expresan en términos de cómo son percibidos (preferencias) y valorados por los pacientes o por la sociedad. Se genera un indicador que combina morbilidad y mortalidad, y las pondera por la preferencia del paciente (utilidad; por ejemplo, años de vida ajustados por calidad).	Razón incremental de costo-utilidad
Costo-beneficio	Costos no establecidos; pueden ser diferentes entre intervenciones. Se enumeran los recursos y se les asignan valores monetarios estandarizados (por ejemplo, en pesos colombianos de 2014).	Se espera que los efectos sean diferentes. Se hace una valoración monetaria de los efectos positivos y negativos sobre el padecimiento y sobre la salud del paciente, y se expresan en cantidades estandarizadas de una moneda (por ejemplo, pesos colombianos de 2014).	Diferencia monetaria entre costos y beneficios

Análisis de minimización de costos

En muchas ocasiones hay que escoger entre tecnologías sanitarias con efectos similares.



En estos casos se emplea una evaluación económica denominada *minimización de costos*, que compara dos o más alternativas, las cuales pueden ser variaciones de una misma tecnología, y que han demostrado experimentalmente que tienen la misma efectividad.

Pero en las cuales es necesario establecer si sus costos son diferentes. Como se sabe que la efectividad es equivalente, la alternativa preferida es la que implica un menor costo.

El investigador debe acudir a alguna fuente capaz de establecer que la seguridad y la efectividad de las alternativas son iguales (por ejemplo, una revisión sistemática) o conducir su evaluación en paralelo con un experimento cuyos resultados demuestren equivalencia de los tratamientos. El investigador debe, entonces, identificar los recursos consumidos y enumerar-

los, y asignarles un valor monetario (costo)[15]. El resultado obtenido (diferencia de costos) corresponde al ahorro teórico que podría lograrse al optar por la alternativa menos costosa; por ejemplo, con la prescripción del mismo principio activo, con igual presentación, pero de la marca que registre el precio más bajo del mercado.

Debe enfatizarse que la comparación es apropiada solo si los medicamentos son ***absolutamente intercambiables***, lo cual significa que deben haberse demostrado bioequivalencia y la misma disponibilidad. El estudio de minimización de costos puede ser visto como un caso especial del análisis de costo-efectividad, donde la evaluación de efectividad demostró equivalencia.

Estudios que contrastan costos y consecuencias

En otros escenarios puede ocurrir que una de las alternativas haya demostrado mayor eficacia y menor costo, de suerte que la escogencia de la mejor opción resulta obvia. En lenguaje de evaluación económica, se afirma que dicha alternativa es ***dominante***. El problema de decisión es complejo cuando ninguna alternativa es dominante. Un tratamiento puede ser más efectivo, pero su administración ser más costosa; o puede tener efectos adversos que sea necesario mitigar, y cuyo tratamiento añade costos, en comparación con otro menos efectivo, pero mejor tolerado o más barato. En tales casos, es necesario balancear de forma objetiva los pros (beneficio, ahorro de recursos) y los contras (costos, efectos adversos). En este tipo de evaluaciones económicas completas, los beneficios pueden ser medidos y valorados en términos de efectividad (clínica), de preferencia del paciente o de la sociedad por los diferentes desenlaces (utilidad) o por el valor monetario atribuido a los efectos positivos o negativos sobre la salud de los pacientes (beneficio).

Estudio de costo efectividad

Un estudio de costo-efectividad compara los costos y las efectividades de dos o más alternativas. El resultado de dicho análisis se denomina *razón de costo-efectividad incremental*. Esta razón se calcula dividiendo la diferencia en costos de la aplicación de las alternativas entre la diferencia en sus efectividades.



El cociente representa la cantidad de dinero adicional que se debe pagar por cada unidad extra de efectividad conseguida al usar la alternativa más efectiva.

Los costos incluyen la valoración monetaria de los recursos consumidos para hacer la intervención, al igual que los costos generados tanto por la morbilidad no evitada como por las complicaciones del tratamiento.

La efectividad corresponde a un indicador único que resuma o capture la efectividad de la intervención[16]. Ese indicador puede ser una consecuencia específica muy importante (muertes evitadas, infartos prevenidos, etc.), pero en la mayoría de los escenarios, tal tipo de indicador no captura el espectro de los efectos sobre la salud. En general, se prefiere un indicador compuesto que resuma morbilidad y mortalidad de una forma estandarizada y razonable en una sola variable; por ejemplo, años de vida saludable (AVISA) ganados.

Este tipo de estudios puede tener dos usos principales: 1) decidir si el balance entre costos y beneficios justifica hacer una intervención dentro de un sistema con recursos escasos (eficiencia distributiva), o 2) comparar diferentes intervenciones para decidir cuál de ellas emplear (***eficiencia productiva***). En ambos casos resulta necesario determinar *a priori* cuánto se estaría dispuesto a pagar por una unidad de los beneficios en salud evaluados. A ese valor se le llama ***umbral de costo-efectividad***[17], debido a que constituye un límite al costo por unidad de beneficio en salud por encima del cual la tecnología es considerada no costo-efectiva.

Hay cinco requisitos básicos para hacer un análisis de costo-efectividad:

- La existencia de más de una alternativa: Es posible que para cierta condición de salud solo haya una alternativa de tratamiento, y en la que sea éticamente incorrecto no hacer nada, pues ello implicaría la muerte del paciente. En ese tipo de casos no deben hacerse evaluaciones de costo-efectividad: tan solo porque una tecnología que puede salvarle la vida a un paciente resulta costosa, no se le puede negar el acceso a ella. Pero no siempre es necesario identificar más de una tecnología para el tratamiento

o el diagnóstico, pues no hacer ninguna intervención puede ser una alternativa, siempre y cuando no signifique la muerte del paciente ni involucre desenlaces graves.

- Que haya diferencias en la efectividad o la seguridad entre las alternativas evaluadas: Si no existe evidencia de que los resultados en salud obtenidos con cada una de las alternativas se diferencian, la decisión únicamente involucraría elementos del consumo de recursos, y deberá conducirse un análisis de minimización de costos.
- Que haya diferencias en el costo de las alternativas: En caso de que el costo de todas las alternativas sea el mismo, la elegida deberá ser la que resulte más efectiva. Dicho balance puede realizarlo el médico tratante, apoyado en la literatura, sin acudir a consideraciones de tipo económico. Sin embargo, suele ser necesario tener una estimación formal de costos, pues para el clínico no siempre es evidente cuáles son los recursos que consume el uso de la alternativa, y muchas veces se necesita estimar, de forma precisa y no sesgada, los llamados ***costos inducidos***, que corresponden a los recursos consumidos para tratar la morbilidad no evitada por un tratamiento, sus complicaciones o las reacciones adversas.
- Incertidumbre sobre si las alternativas más efectivas son más costosas: Cuando haya certeza de que la alternativa más efectiva (incluyendo todos los costos de tecnologías complementarias y el tratamiento de los posibles efectos adversos) es, a su vez la menos costosa, esta debe ser considerada la alternativa dominante (como ya se mencionó), y no se requiere hacer una evaluación económica.
- Que haya variabilidad en la práctica clínica: Las evaluaciones económicas de tecnologías sanitarias constituyen información útil en la toma de decisiones solo en los casos donde hay incertidumbre. Si, dadas las condiciones de salud de un paciente, el criterio médico indica que solo existe una alternativa, o si para un estado de salud, pese a la existencia de varias opciones de tratamiento o diagnóstico, todos los expertos clínicos consideran que es claro cuál de ellas elegir, no es necesaria ni adecuada la realización de un análisis de costo-efectividad.

Como ya se mencionó, el resultado principal de un análisis de costo-efectividad es la ***razón de costo-efectividad incremental***, también denominada **ICER** por las siglas en inglés de *incremental cost effectiveness ratio*, y que es la forma más adecuada de evaluar la ***eficiencia productiva***; o sea, la alternativa más eficiente para manejar un problema específico en salud.

La razón de costo-efectividad incremental muestra el costo en el que se incurriría por cada unidad adicional de beneficio en salud, en comparación con el uso de otra tecnología. En este caso, el numerador corresponde a la diferencia entre los costos totales de la intervención evaluada y los de la tecnología que sirve de comparador (***costo incremental***); a su vez, el denominador será la distancia entre los resultados en salud de la tecnología evaluada y la que se emplea para la comparación (***efectividad incremental***).

EJEMPLO

Se evalúa el tratamiento X, en comparación con la opción Y. El costo del primero es de $1 000 000 por cada 1000 pacientes, mientras que el del segundo es de $800 000 por cada 1000 pacientes; al acudir a la tecnología X, se extiende, en promedio, 2 años la esperanza de vida del paciente, pero si se utiliza Y, el incremento es de un año. El costo incremental sería de $200 000, y la efectividad incremental, de 1000 años (un año por cada paciente en una cohorte de 1000 individuos). En consecuencia, la razón de costo-efectividad incremental será de $200, y se concluiría que cada año de vida ganado con la tecnología X, adicionalmente a los obtenidos con la alternativa Y, tendría un costo de $200.

También se puede calcular para cada alternativa una ***razón de costo-efectividad***, que es, simplemente, el cociente entre el costo y el efecto de cada alternativa.

Esta razón es útil para comparar tecnologías en salud empleadas en problemas o patologías diferentes que compiten por recursos de un presupuesto (***eficiencia distributiva***), por ejemplo, cuando se decide cuántos recursos asignar a un programa de vacunación, comparados con programas de planificación familiar, prevención de accidentes de tránsito o programas de trasplante de órganos.

EJEMPLO

(Razón de costo-efectividad simple): Si el costo total del uso de la intervención X, que es una vacuna para prevenir una enfermedad infecciosa grave es de $1 000 000 por cada 1000 pacientes, y con ella se logra incrementar la esperanza de vida de cada individuo en 20 años, la razón de costo-efectividad será de $50, lo cual significaría que cada año de vida ganado atribuible a la intervención tiene un costo de $50. Esta intervención se compara favorablemente con una hipotética profilaxis antibiótica para pacientes con antecedente de pielonefritis, que podría costar $100 000 000 por cada 1000 pacientes tratados por un año, y prolongaría en 2 años la expectativa de vida de cada paciente: $50 000 por año de vida ganado. Tal tipo de información le resulta útil al formulador de políticas que debe asignar los recursos del presupuesto a diferentes programas o tecnologías en salud (***eficiencia distributiva***).

Vale la pena advertir al lector que no es adecuado utilizar razones de costo-efectividad para evaluar la eficiencia productiva. En ese caso se debe usar la razón de costo-efectividad incremental. Cuando se analizan alternativas con el fin de escoger la más eficiente para tratar una patología específica en un grupo poblacional determinado, la comparación de razones de costo-efectividad simples es engañosa, pues subestima el incremento en costo, al pasar de una alternativa menos efectiva (y menos costosa) a una más efectiva, y puede llevar a elegir alternativas ineficientes.

EJEMPLO

(Eficiencia productiva): Cuando se usa la estatina A para la prevención secundaria de infarto agudo de miocardio, se encuentra que para tratar a 100 pacientes se gastan $6 000 000 por 5 años, y en esos 5 años se previenen 10 infartos. Si se utiliza la estatina B (más efectiva), se previenen 20 infartos por cada 100 pacientes por 5 años de tratamiento, pero el costo se eleva a $18 000 000. La razón de costo-efectividad de la estatina A es 6 000 000 / 10 = $600 000 por infarto prevenido. La de la estatina B es $18 000 000 / 20 = $900 000. Aparentemente, usar B cuesta solo $300 000 más por evento evitado, y previene el doble de eventos que la estatina A. Lo que este raciocinio no tiene en cuenta es que si se hubiese utilizado la estatina A, ya se habrían prevenido 10 infartos. Al usar la estatina B no se evitan 20 infartos: se evitan 10 infartos *adicionales a los que habría evitado la A*. El uso de la razón de costo-efectividad incremental da una mejor idea de las consecuencias económicas de pasar de la estatina A a la estatina B:

$$\frac{\text{Costo de B (\$18 000 000)} - \text{Costo de A (\$6 000 000)}}{\text{Efecto de B (20)} - \text{Efecto de A (10)}}$$

Esta razón incremental es de $1 200 000 adicionales que se deben pagar por cada infarto adicional prevenido, lo cual es 4 veces más que lo que sugería la comparación de las razones simples de costo-efectividad de cada alternativa.

Las razones de costo-efectividad, se trate de simples o de incrementales, son, en esencia, la división entre el costo total (o diferencial) del uso de las tecnologías evaluadas y la totalidad (o la diferencia) de los beneficios en salud obtenidos. De ahí la necesidad de representar los resultados en salud con una única medición escalar que permita efectuar la suma y obtener el divisor. El cociente resultante será el costo de cada unidad de beneficio en salud.

El supuesto tras la razón de costo-efectividad incremental es que la tecnología evaluada es más costosa y más efectiva que la alternativa empleada para la comparación.

Si no es así, el valor obtenido debe interpretarse cuidadosamente. En caso de que el numerador sea negativo y el denominador positivo, la opción evaluada es menos costosa y más efectiva, por lo cual sería una estrategia dominante, y, posiblemente, se hizo la evaluación porque solo hasta incluir los costos de las tecnologías complementarias, así como los del tratamiento de los posibles efectos adversos, y llevar a cabo una simulación, fue posible establecer la diferencia a favor en el costo de la intervención más efectiva. Si el numerador es positivo y el denominador negativo, se evaluó una tecnología que resultó más costosa, pero menos efectiva, con lo cual el comparador sería la alternativa dominante. Por último, cuando dividendo y divisor son negativos, la opción evaluada es menos costosa y menos efectiva; en tal caso, el valor correspondería a la razón de costo-efectividad incremental del comparador, y no de la tecnología evaluada.

Como ya se mencionó, la evaluación de costo-efectividad puede hacerse en paralelo con un experimento que evalúe la efectividad y la seguridad, o en un modelo. Cuando se necesita estimar la eficiencia de distintas tecnologías para un problema específico en una población determinada, extrapolar la evaluación económica de un experimento puede ser inadecuado. En estos casos (y debido a que tanto para el cálculo de los costos como para la estimación de los resultados en salud deben tenerse en cuenta, además del tratamiento, las tecnologías complementarias para un correcto uso de la tecnología sanitaria, los posibles desenlaces y el tratamiento de los efectos adversos, así como los riesgos basales y los perfiles pronósticos de la población de interés) es preferible elaborar un modelo que permita simular una cohorte hipotética (pero que, además, capture las características de interés de la población blanco) por cada alternativa evaluada.

Todas las cohortes simuladas deben tener las mismas condiciones iniciales, de tal forma que se establezca una comparación justa entre las tecnologías.

Un *modelo* es cualquier abstracción que emplea supuestos para simplificar la realidad, y deja solo los elementos que pueden afectar de manera considerable el resultado final.

Por tal motivo, en su construcción se debe involucrar a expertos clínicos que avalen los supuestos empleados y se evite un exceso de simplificación que invalide los resultados.

En general, hay tres tipos de modelo empleados en las evaluaciones de costo-efectividad: 1) los árboles de decisión, 2) los modelos de Markov y 3) la simulación de eventos discretos. No es el propósito del presente texto detallar las metodologías de modelación; sin embargo, es importante señalar que todas ellas se sirven de la epidemiología y utilizan los resultados de metaanálisis, experimentos clínicos o estudios de cohortes para establecer las probabilidades empleadas en la simulación.

Estudios de costo-utilidad

El modelo biomédico tradicional se basa en la evidencia empírica de carácter científico, y emplea indicadores cuantitativos clásicos, como la mortalidad o los años de vida ganados. Sin embargo, en los años noventa del siglo **xx** empezó el auge de adicionar al conjunto de información empleado en la toma de decisiones en salud la percepción del paciente sobre la calidad de vida derivada del estado de salud resultante del uso de cada posible alternativa.

El concepto de calidad de vida es, en esencia, la valoración subjetiva que el individuo hace sobre el nivel de bienestar que percibe, de acuerdo con sus propias condiciones físicas, mentales, emocionales y sociales en determinado momento.

En consecuencia, se ve afectado por factores del medio ambiente, por aspectos socioculturales y por la manera como el paciente se relaciona con su entorno.

Tener en cuenta la calidad de vida del paciente enriquece el proceso de decisión al punto de que puede ser el factor determinante. Por ejemplo, pese a que el médico pueda preferir el tratamiento que reporte el mayor impacto sobre la ***esperanza de vida*** (la percepción del paciente sobre la manera como vivirá esos años) puede hacer que cambie la decisión, pues algunas circunstancias (las manifestaciones de la enfermedad o de los medicamentos, los síntomas o las limitaciones) pueden hacer que el individuo prefiera la muerte, en vez de extender su vida en condiciones que considera adversas.

Los estudios de costo-utilidad son una evolución de los estudios de costo-efectividad, en tanto emplean una medición de la calidad de vida en lugar de la mera cuantificación de un desenlace de efectividad. Por tal razón, comparten todas sus características, exceptuando que en vez de, simplemente, expresar los resultados en salud en términos de las necesidades o las discapacidades evaluadas por un observador (clínico), se incorpora la preferencia (subjetiva) que el paciente o la sociedad expresan por estados discretos y específicos de salud, medidos por un instrumento estandarizado (objetiva) como, por ejemplo, los inventarios generales o específicos de estados de salud y funcionalidad, conocidos globalmente como ***escalas de calidad de vida*** *relacionadas con salud*.

La definición de ***utilidad*** en teoría de decisiones y evaluaciones económicas en salud

proviene de los trabajos sobre teoría de juegos de von Neumann y Morgestern, en 1944[18]. Ambos autores definieron *utilidad* como la preferencia explícita y cuantitativa, expresada por un tomador de decisiones, frente a un desenlace específico: dicha preferencia es medida en una escala de intervalo (numérica continua con un cero arbitrario) y en presencia de incertidumbre (riesgo o probabilidad); es decir, sin que el tomador de decisión tenga certeza sobre la ocurrencia de los desenlaces posibles. Desafortunadamente, von Neumann y Morgestern denominaron a esa preferencia, medida cardinalmente, "utilidad", lo que se ha prestado para confusiones.

En la teoría microeconómica contemporánea, se entiende por utilidad el grado de bienestar o de satisfacción de un consumidor, y se usa como parte de la teoría del consumo. Las formulaciones hechas durante el siglo XIX sobre teoría del consumo postularon una utilidad medida con números cardinales, en presencia de certidumbre de consumo del bien[10]. La visión microeconómica contemporánea, a partir de Vilfredo Pareto, al comienzo del siglo XX, y refinada posteriormente[19], se fundamenta en el ordenamiento que hace el consumidor de sus preferencias, incluyendo el riesgo. Los estudiantes de economía aprenden que la *utilidad cardinal*, medida en presencia de certidumbre, es subjetiva, no mensurable, y que, incluso, puede no existir.

En contraste, en teoría microeconómica se define la utilidad como el bienestar que el individuo obtiene como resultado de consumir un determinado bien o servicio[20,21], y medido en una escala ordinal. En consecuencia, muchos economistas creen que las mediciones de la calidad de vida empleadas en estudios de costo-utilidad son valoraciones subjetivas del bienestar asociado a un estado de salud, y que, por ello, a este tipo de análisis se le da, inapropiadamente, el nombre de costo-utilidad, pues se entiende que las mediciones de calidad de vida empleadas no necesariamente cumplen con las condiciones de la teoría de la utilidad en economía[22]. Esta afirmación es errónea, tanto histórica como conceptualmente. Las utilidades de von Neumann y Morgestern tienen una sólida base teórica, y fundamentan la teoría contemporánea del análisis de decisiones (ampliamente utilizada y validada en negocios, gobierno, atención en salud, inversión económica, estrategia y táctica militar y muchos otros campos), con demostrado éxito por muchas décadas, pues son teórica y empíricamente estables y diferentes del concepto microeconómico de utilidad.

Existen diferentes unidades de medida para la calidad de vida, y la más utilizada es el denominado *Quality Adjusted Life Years* (QALY), o en español, años de vida ajustados por calidad (AVAC), concepto basado en un concepto muy sencillo: cada año es ponderado de acuerdo con la preferencia que el paciente (o la sociedad) expresa por su nivel de calidad de vida relacionada con salud; la preferencia se cuantifica entre 0-1, donde 1 es un estado teórico de perfecta salud, y 0, la muerte.



Un QALY equivale a un año vivido en perfecta salud[22,23].

Por ejemplo, si un paciente sobrevive 8 años en un estado del cual se considera que es tan solo el 25 % de la salud perfecta, se diría que el resultado obtenido es de 2 QALY.

Aunque el concepto es siempre el mismo, hay diferentes maneras de obtener las preferencias con las que se calculan los QALY. El más común es, a su vez, el más sencillo, y consiste en el uso de la escala visual análoga (EVA). En esta, cada persona indica, en una escala de 0-1, o de 0-100, para facilitar la comprensión del instrumento, la valoración o el puntaje que le otorga a cierto estado de salud. Como esta asignación de preferencia solo tiene en cuenta el nivel de calidad de la vida relacionada con la salud, sin considerar la probabilidad de llegar a dicho estado (certidumbre), el método logra que el sujeto exprese un tipo de preferencia conocido como *valor*.

En contraste, para obtener utilidades de von Newmann y Morgestern se deben buscar las preferencias expresadas por los pacientes *en presencia de incertidumbre*. Los métodos buscan que el paciente elija entre alternativas con probabilidades diferentes y consecuencias o estados de salud distintos, y no una simple valoración de la preferencia por un estado cierto. Dentro de esas técnicas se encuentran el *intercambio temporal*, o negociación de tiempos (TTO, por las siglas en inglés de *time trade-off*), también llamado *compensación temporal*,

y el cual consiste en lograr que el individuo exprese el número mínimo de años que estaría dispuesto a vivir con la condición de salud que se quiere valorar, en vez de contar únicamente con un año de perfecta salud. Si se denomina α al número indicado por el individuo, cada año de vida de esa condición en salud equivaldría a $1/\alpha$ QALY.

EJEMPLO

Para encontrar la utilidad que un sujeto le asigna a vivir con una resección de la laringe y una traqueostomía permanente por causa de la extirpación radical de un cáncer de laringe, se le proponen al sujeto dos escenarios: 1) vivir 20 años curado de su cáncer, pero sin laringe y respirando a través de una traqueostomía, o 2) que se le practique una cirugía conservadora, que solo extirpa el tumor, y el paciente conserve su laringe, pero solo para sobrevivir un año. Supongamos que el sujeto declara que prefiere vivir 20 años, aún sin laringe. Entonces se modifica el escenario diciéndole que con la extirpación del tumor y preservando la laringe sobrevivirá 2 años. Si el paciente aún prefiere vivir esos 20 años sin laringe, se eleva la supervivencia en el caso de la cirugía conservadora. Eventualmente, esa persona, cuando le solicitan que escoja entre vivir 20 años sin laringe y sin cáncer, o 5 con laringe y fallecer a causa del tumor, el sujeto es incapaz de elegir entre los 2 escenarios. En ese momento, ha expresado cuantitativamente y en presencia de incertidumbre (utilidad) que, en su opinión, vivir 5 años en plena salud es equivalente a vivir 20 años sin cáncer, pero con secuelas graves (con traqueostomía y sin laringe). Al dividir los 5 años con perfecta salud entre los 20 años con traqueostomía, se comprende que un año con traqueostomía es solo el 25 % o una cuarta parte de un año en perfecta salud. Dicho de otro modo, 20 años con traqueostomía representan 5 QUALY.

Otra alternativa es la ***apuesta estándar***. En ella, el individuo debe decidir bajo qué nivel de incertidumbre aceptaría una intervención hipotética que, con una probabilidad β, lo curaría de la enfermedad, y a la cual se le quiere estimar el nivel de calidad de vida asociado, pero cuenta con una probabilidad $1-\beta$ de causarle la muerte. El valor mínimo de β, por el que el individuo aceptaría someterse al tratamiento, refleja la valoración en términos de la calidad de vida asociada a la enfermedad.

También es posible ponderar los años de vida según la discapacidad presente; este es el origen de los ***disability adjusted life years*** (DALY), o años de vida ajustados por discapacidad (AVAD). En este caso, nuevamente, el ponderador será un número real entre cero y uno, pero su significado varía: cero corresponde a la ausencia de discapacidad, y uno, a la discapacidad total. Un DALY es un año perdido de vida saludable[24,25]. Como los años de vida saludable pueden perderse a causa de la muerte o de la discapacidad resultante, el total de los DALY es igual a la suma de los años perdidos por muerte prematura y los años perdidos por discapacidad. El hecho de que los DALY reflejen un resultado negativo hace que sean empleados particularmente en los estudios de carga de la enfermedad. En realidad, los DALY no representan utilidades, pues no se ponderan por las preferencias de los pacientes o de la sociedad en presencia de incertidumbre, sino que su ponderación proviene de valorar de manera "objetiva" necesidades y limitaciones que los expertos en salud asignan al estado de salud del sujeto.

EJEMPLO

Unos gemelos sufren un accidente idéntico, a raíz del cual se les fractura el brazo derecho. Mientras que el nivel de incapacidad, juzgado de forma estandarizada por un clínico, es el mismo para ambos hermanos (la misma cantidad de DALY asociados a la fractura), la preferencia expresada por el dolor y la limitación causada por la fractura pueden ser valoradas de forma diferente, aun en el caso de gemelos. Si se computan los QALY asociados a dicho estado de salud alterado, estos pueden (y suelen) ser diferentes, mientras que los DALY medidos serán idénticos.

El cálculo de equivalentes de años de vida saludable ponderados por discapacidad es particularmente apropiado como medida única de resumen de la efectividad y la seguridad de intervenciones, y se usa, tal como ya se mencionó, en análisis de costo-efectividad. Por contraste, en los análisis de costo-utilidad es bastante común que la utilidad se mida en QALY.

Estudios de costo-beneficio

Análisis de costo-beneficio

Si se quiere evaluar el beneficio de una intervención en salud, puede usarse un diseño que

compara alternativas que compitan, en términos monetarios, tanto para los costos como para los beneficios. Este diseño tiene gran utilidad para evaluar la eficiencia en asignación de recursos, y, a diferencia de los otros diseños, los cuales se centran en eficiencias relativas, puede utilizarse para evaluar programas de manera más absoluta. De hecho, tal enfoque es el que mejor satisface los supuestos de una evaluación económica, pues expresa en términos equivalentes (monetarios) tanto los recursos consumidos (costos) como los beneficios obtenidos. También es, en términos de salud, el que más inquietudes genera, por la asignación de términos monetarios a los beneficios obtenidos (un año de vida salvado, por ejemplo). El diseño puede responder a la pregunta de si la intervención vale la pena por sí misma, o si los beneficios son mayores que los costos.

Una ventaja de este diseño es la posibilidad de comparar el impacto económico neto con otras actividades muy diferentes: por ejemplo, cuánto se beneficia (gana) la sociedad con la adopción de un nuevo medicamento y cuánto se beneficia construyendo un nuevo hospital o un centro comunitario.

Por supuesto, la dificultad principal estriba en la conversión de beneficios de un programa de atención en salud a valores monetarios. La teoría económica prescribe que el valor monetario de los beneficios debería estimarse a partir de su valor en el mercado. Como no hay una forma directa de estimar el valor de mercado de los desenlaces en salud, se emplean alternativas indirectas para intentar asignar un valor apropiado. Las dos principales soluciones para este difícil paso son:

- Capital humano: **Permite asignar valor al medir** cómo la intervención le ayuda al paciente a recuperar (o a aumentar) su capacidad productiva, y que, a su vez, puede medirse mediante el salario o un subrogado de este. Si bien es un proceso aproximado, permite un buen acercamiento, aunque no sea aplicable para niños ni para ancianos.
- Disposición a pagar: **Se centra en la máxima** cantidad que una persona estaría dispuesta a pagar para tener la utilidad que pueda ganarse de la intervención. Aunque es también subjetiva, el hecho de que el paciente deba decidir cuánto quiere pagar de sus ingresos le da un valor especial. Se le presentan al individuo escenarios hipotéticos, para que él tome la decisión de cuánto estaría dispuesto a dar para obtener un beneficio, o para evitar un costo. Esta es un área de dificultad, y se han desarrollado muchas técnicas para tal valoración.

El estudio, por su parte, consta de tres fases principales:

- Definición de propósito: **Se definen la intervención** y el beneficio, y en algunos casos, una alternativa para el gasto (o, al menos, el costo) de oportunidad.
- Costos: **Se hace una determinación detallada** de los costos (directos, indirectos e intangibles), y se tiene en cuenta el cambio futuro del dinero (inflación o deflación).
- Beneficios: **El aspecto más complejo del diseño**; se logra con las estrategias comentadas, entre otras: capital humano y disposición a pagar. Debe considerarse que puede haber un beneficio claro (control de una enfermedad), pero también puede ser la mejoría del nivel general de salud o la prevención de complicaciones o de enfermedades.
- Comparaciones y conclusión: **Se comparan los** costos asociados con los beneficios potenciales y los costos de la intervención; también se comparan con la intervención alternativa, si existe, y se hace el análisis del costo de oportunidad, si no lo había y es apropiado hacerlo.

COSTOS EN SALUD

Definición

Dependiendo del contexto, el costo puede tener múltiples definiciones.



Para el caso de las evaluaciones económicas de tecnologías sanitarias el costo es el valor monetario de utilizar un recurso como parte de cualquier tipo de intervención que permita obtener un bien o un servicio en salud[26].

La perspectiva en la estimación de los costos

En cada sistema de salud la distribución del peso presupuestal sobre los actores (paciente, cuidador, prestador y asegurador) es diferente.

Por esta razón el primer elemento que debe determinarse en el proceso de estimación de costos es la *perspectiva de costeo*[27]; es decir, quién asumirá el costo que se está estimando.

Esta diferenciación es importante también porque lo que para alguno de los actores es un costo, para otro puede ser el precio de venta. Por ejemplo, en Colombia, un paciente del régimen contributivo (paga regularmente para recibir atención en salud, y aporta para que la reciban quienes no pueden pagar), diagnosticado con leucemia mieloide aguda, que deba ser llevado a alotrasplante con células madre compatibles, no tendrá que pagar por el trasplante o por los estudios de histocompatibilidad; tampoco, por los procedimientos, los medicamentos ni los insumos necesarios para el tratamiento de los posibles eventos adversos. En consecuencia, si se quiere desarrollar un estudio de costos para el trasplante con la perspectiva del paciente, no deberá incluirse el valor del procedimiento ni el tratamiento de los efectos adversos: únicamente se tendrán en cuenta los pagos o los gastos en los cuales deba incurrir el paciente al no estar cubiertos por su plan de salud. En caso de que la perspectiva sea la de un paciente no adscrito a algún sistema de aseguramiento personal o colectivo, sí deberán ser incluidos todos los cobros hechos por la institución prestadora.

El ejercicio de costeo llevado a cabo con la perspectiva del asegurador, también llamado *tercero pagador*, tendría en cuenta los costos del trasplante según el precio y las condiciones de negociación que haya acordado con la institución prestadora de servicios de salud (IPS). Es posible que se pacte un valor fijo por cada trasplante, independientemente del consumo de recursos de cada caso en particular; de esta manera, se traslada el riesgo, y, sin importar que en el tratamiento completo se consuman recursos por un valor diferente del acordado, el asegurador únicamente pagará el valor pactado. En consecuencia, ese será el costo para incluir en el análisis, no el consumo real de recursos.

Los valores pactados entre aseguradores e IPS usualmente son consignados en manuales tarifarios, y estos son empleados como base para negociaciones futuras. Cabe recordar que las tarifas cobradas por cada servicio pueden no reflejar de forma absoluta ni relativa el costo de producción. A menudo, las instituciones hacen subsidios cruzados a la hora de fijar tarifas para ventas de servicios. Por ejemplo, se hace un descuento en el cobro de un procedimiento quirúrgico, pero se compensa cobrándole al tercero pagador tarifas mayores en otros servicios prestados al mismo paciente, tales como medicamentos o pruebas de laboratorio.

Debido a lo anterior, cuando el estudio se hace con la perspectiva de la IPS, no es adecuado recurrir a los valores acordados con los aseguradores, ya que los precios de venta (tarifas) no necesariamente reflejan los costos. En su lugar, serán empleados los valores a los cuales son adquiridos cada uno de los elementos necesarios para prestar el servicio: por ejemplo, honorarios del personal médico, pago por medicamentos e insumos, servicios públicos, etc.

Finalmente, si se quiere cuantificar el costo incluyendo a todos los posibles actores, debe emplearse la perspectiva de la sociedad en su conjunto, lo que implica determinar todos los posibles costos asociados a una determinada intervención, sin importar quién deba asumirlos.

Clasificación

En el ámbito de salud, hay diferentes maneras de clasificar los costos **(tabla 23.4)**. Los criterios más comunes son el ***contable***, la ***relación con la intervención***, la ***posibilidad de verificación*** y la ***probabilidad de ocurrencia***. La clasificación empleada no determina cuáles costos estimar, sino la manera como serán agrupados para el análisis, por lo cual la elección depende del objetivo que quiera cumplirse.

- Contablemente, los costos pueden clasificarse como ***fijos*** y ***variables***. Los primeros no dependen del número de individuos atendidos, pero los segundos, sí. Por ejemplo, si un hospital no es propietario de sus instalaciones y debe pagar un canon de arrendamiento, dicho costo se considera fijo, debido a que, sin importar cuántos pacientes acudan al servicio, el costo siempre será el mismo. Mientras, del número de hemogramas que se hagan dependerá la cantidad de jeringas, de tubos y de gasas empleadas, y, en consecuencia, también el valor monetario total que deba pagarse por ellos.
- Si los costos se clasifican según la relación con la intervención, se tendrán ***costos médicos directos***[28], ***costos directos no médicos*** y ***costos indirectos***[29]. Los costos médicos directos son los referentes a los honorarios

Tabla 23.4 Clasificación de costos

• Criterio contable.
• Costos fijos.
• Costos variables.
• Relación con la intervención[1].
• Costos del sector salud o costos médicos directos.
• Costos del paciente o su familia, y costos de otros sectores diferentes de salud (o costos directos no médicos).
• Costos de productividad perdida, o costos indirectos.
• Método del capital humano.
• Disponibilidad para pagar.
• Fricción de costos.
• Posibilidad de verificación.
• Probabilidad de ocurrencia.

[1] Drummond[10] propone (y ello ha sido ampliamente aceptado en la literatura de evaluación económica anglosajona) no distinguir entre costos "directos" e "indirectos", ya que hacerlo se presta a confusión. En contabilidad analítica, los costos indirectos se refieren a los costos del uso de infraestructura y de recursos compartidos, que se distribuyen proporcionalmente entre distintos procesos productivos que comparten recursos y servicios (por ejemplo, arriendos, servicios generales y públicos, celaduría, etc.). Por eso, propone hablar de costos de producción del sector salud, del paciente y de otros sectores ("costos directos médicos y no médicos") y costos causados por la productividad perdida asociada a la enfermedad o a su tratamiento, en vez de llamarlos costos indirectos.

del personal médico, los medicamentos y los insumos que se consumen durante el procedimiento. Los costos directos no médicos son todos los pagos por elementos o actividades necesarios para la intervención, pero que no hacen parte de ella: por ejemplo, el traslado del paciente hasta el centro médico donde pueda practicarse el procedimiento. Finalmente, los costos indirectos son el valor monetario del tiempo que el paciente o sus cuidadores deben destinar para llevar a cabo el procedimiento y para la recuperación.

Para valorar los costos indirectos hay tres metodologías. La primera de ellas es el ***método de capital humano***, consistente en tener ***a priori*** el valor hora según la profesión, o el producto interno bruto per cápita, y multiplicarlo por el tiempo de incapacidad. La segunda es la ***disponibilidad para pagar***, que, de manera consultiva, construye cuánto está dispuesto a pagar el empleador por el recurso humano comprometido. La tercera, es conocida como ***fricción de costos***, la cual tiene en cuenta el costo de la capacitación requerido por la persona que reemplazará a quien se encuentra incapacitado.

En general, podrá hallarse evidencia física o documental de la realización de actividades y del consumo de recursos asociados. En esos casos se dice que se trata de ***costos tangibles***; sin embargo, hay afecciones sobre el bienestar del paciente cuyo valor monetario también puede ser estimado sin que pueda registrarse una evidencia objetiva, por lo que son llamados *costos intangibles*, generalmente asociados al dolor y a la disposición para pagar con tal de no padecerlo.

Finalmente, la probabilidad de ocurrencia de los eventos que generan un consumo de recursos (y, en consecuencia, un costo) permite separar los costos evitables de los no evitables. Los ***costos evitables*** son los costos en los cuales no se va a incurrir si se hace una serie de intervenciones que no necesariamente son médicas. Un ejemplo es la incidencia del trauma por accidente de tránsito en Bogotá, el cual, gracias a la implementación de políticas como el uso obligatorio del cinturón de seguridad, y del casco para motociclistas, disminuyó considerablemente; en consecuencia, también se redujo la cantidad de recursos destinados a atender este tipo de eventos.

Los ***costos no evitables*** existen con motivo de la misma funcionalidad o el desempeño del ser humano y de las acciones que, según el criterio médico, deben llevarse a cabo para prevenir enfer-

medades, mejorar la calidad de salud de la población y no poner en peligro la vida o la integridad; por ejemplo, las campañas de vacunación.

Secuencia de costeo

Lo primero que se hace respecto a esta secuencia es identificar los eventos generadores de costo y los recursos utilizados en cada uno de ellos. Ejemplo: una prueba de tiempo parcial de tromboplastina (PTT, por las siglas en inglés de *partial thromboplastin time*), una consulta o una cirugía. Los eventos generadores de costo son las actividades que se hacen durante las intervenciones en salud (las de cualquier orden). Después de identificarlos, los eventos generadores deben ser *cuantificados*; es decir, debe establecerse cuántas veces ocurre el evento durante la intervención (y por ello solo expertos clínicos pueden hacerlo adecuadamente). El tercer y último paso consiste en valorar el evento, calcular cuánto cuesta de acuerdo con la perspectiva desde la cual se esté valorando.

Métodos

La metodología de costeo se refiere a la manera como serán identificados y agrupados los eventos generadores y los recursos empleados. Cada una de las metodologías tiene un nivel de complejidad directamente proporcional a la precisión con la que se calculan los costos. Ordenadas de la más compleja y precisa a las más sencilla e inexacta, están: 1) el microcosteo, 2) el costeo basado en actividad, 3) el caso tipo, 4) los grupos de diagnósticos relacionados (GDR), 5) la factura emitida, 6) el ajuste de costo y 7) el promedio generador de costo[30] **(tabla 23.5)**.

- Microcosteo: *Identificación*, con el mayor nivel de detalle posible, de todos los eventos generadores de costo en las intervenciones, sin hacer ningún tipo de agrupamiento; por ello, es la metodología más precisa.
- Costeo basado en actividad: Se realiza por centros de costo (cirugía, unidad de cuidados intensivos —UCI—, ginecoobstetricia), y agrupa eventos generadores en una sola actividad. Usualmente, corresponde al precio al que la entidad prestadora ofrece el servicio, por lo cual los valores pueden ser tomados de manuales tarifarios. El elemento crucial en esta metodología es la adecuada identificación de las actividades. Es posible que haya en los manuales definiciones similares u homónimos. Un ejemplo es la PCR, que puede corresponder a proteína C reactiva o a una reacción de polimerasa en cadena. Son dos eventos distintos, con valores diferentes, pero que en el lenguaje médico son denominados con la misma sigla.
- Caso tipo: Consiste en construir un caso representativo que tenga en cuenta lo observado más a menudo en la práctica clínica, o adoptar las intervenciones recomendadas en guías de práctica clínica (GPC). La diferencia con los anteriores radica en que el caso tipo es construido mediante la orientación de los expertos, quienes se basan, a su vez, en los datos de la práctica real y excluyen los eventos poco comunes.
- Grupos de diagnósticos relacionados (GDR): los GDR agrupan una serie de procedimientos preestablecidos que, se supone, le serán practicados a cada paciente con determinado diagnóstico.
- Ajuste de costo: En esta metodología se transforma el costo calculado para otras latitudes expresándolo en la moneda legal del país donde se está realizando el estudio, y empleando la correspondiente tasa de cambio. Según ello, se asume que la tecnología se aplica de la misma forma, y que el mercado es similar al del país en que se hizo el estudio original.
- Factura emitida: Se refiere al documento contable emitido por la institución al egreso del paciente, y puede ser muy útil solo si hay de por medio un sistema de registro válido, confiable y oportuno. Para establecer la calidad de los registros, es necesario compararlos con la historia clínica.

Tabla 23.5 Metodologías para costeo

• Microcosteo
• Costeo basado en actividad
• Caso tipo
• Grupos de diagnósticos relacionados (GDR)
• Factura emitida
• Ajuste de costo
• Promedio generador de costo

- Promedio generador de costo: Se trata de extraer información de bases de datos y dividir el costo total registrado para cierta intervención por el número de pacientes a quienes se les practicó. Nuevamente, la validez dependerá de la calidad de información.

Referencias

1. Pharoah PDP, Hollingworth W. Cost effectiveness of lowering cholesterol concentration with statins in patients with and without pre-existing coronary heart disease: life table method applied to health authority population. BMJ. 1996;312:1443-8.
2. Cuevas H. Fundamentos de la economía de mercado. Bogotá: Universidad Externado de Colombia; 2004.
3. Yandle B. Emerging market in air pollution rights. Regulation. 1978;2:21.
4. Polesetsky M. Will a market in air pollution clean the Nation's Dirtiest Air-A study of the South Coast air quality management District's Regional clean air incentives market. Ecol LQ. 1995;22:359.
5. Brennan TA, Rothman DJ, Blank L, et al. Health industry practices that create conflicts of interest: A policy proposal for Academic Medical Centers. JAMA. 2006;295:429.
6. McKeeargue M. Budget crises, health, and social welfare programmes. BMJ. 2010;341:77.
7. Barón-Leguizamón G. Gasto nacional en salud de Colombia 1993-2003: composición y tendencias. Rev Salud Pública Bogotá. 2007;9:167-79.
8. Ooms G, Van Damme W, Baker BK, et al. The 'diagonal'approach to Global Fund financing: a cure for the broader malaise of health systems. Glob Health. 2008;4.
9. Parada L, Taborda-Restrepo A, Chicaíza-Becerra L. Evaluación económica de tecnología sanitaria y toma de decisiones en salud. Coyunt Económica. 2013;XLIII:81-95.
10. Drummond M, Sculpher M, O'Brien B, et al. Methods for the economic evaluation of health care programmes. Oxford: Oxford Medical Publications; 2005.
11. Zárate V. Evaluaciones económicas en salud: Conceptos básicos y clasificación. Rev Médica Chile [internet]. septiembre de 2010 [citado 2014 sep. 1]. Disponible en: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872010001000007&lng=en&nrm=iso&tlng=en
12. Espinoza SM. Evaluaciones económicas en salud en Chile: hacia un marco conceptual más amplio. Rev Médica Chile. 2011;139:816-9.
13. Mejía A, Moreno M. Evaluación económica para la toma de decisiones en salud en Colombia. Monit Estratég - Supersalud. 2014;5:70-5.
14. Darbà J. Los modelos de decisiones en la evaluación económica de tecnologías sanitarias: Calidad y buenas prácticas metodológicas. PharmacoEconomics Span Res Artic. 2013;3:31-8.
15. Luengo-Fernández R, Gray A, Mehta Z, et al. Acute costs of stroke in the UK National Health Service in 2002-2004. Value Health. 2004;7:692.
16. Mandelblatt JS, Fryback DG, Weinstein MC, et al. Assessing the effectiveness of health interventions for cost-effectiveness analysis. J Gen Intern Med. 1997;12:551-8.
17. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the NICE cost effectiveness threshold [internet]. Centre for Health Economics, University of York; 2013. [citado 2014 sep. 1]. Disponible en: http://ideas.repec.org/p/chy/respap/81cherp.html
18. Von Neumann J, Morgenstern O. Theory of games and economic behavior (60th Anniversary Commemorative Edition). New Jersey: Princeton university press; 2007.
19. Arrow KJ, Debreu G. Existence of an equilibrium for a competitive economy. Econom J Econom Soc. 1954;265-90.
20. Varian HR. Microeconomía intermedia: Un enfoque actual [Internet]. Antoni Bosch Editor; 2003. [citado 2014 sep. 1]. Disponible en: http://books.google.com.co/books?id=-rWfisFu618C
21. Nicholson W. Teoría microeconómica: Principios básicos y ampliaciones [internet]. Cengage Learning Latin America; 2006. [citado 2014 sep. 1]. Disponible en: http://books.google.com.co/books?id=D-tRgvKyrikC
22. Chicaíza L, García Molina M, Romano G. Años de vida ajustados por calidad: evolución, fundamentos e implicaciones (Quality Adjusted Life Years: Evolution, Fundamentals and Implications). SSRN Electron J [Internet]. 2013 [citado 2014 sep. 1]. Disponible en: http://www.ssrn.com/abstract=2261511
23. Nord E. Cost-value analysis in health care: making sense out of QALYs. New York: Cambridge University Press; 1999.
24. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic

analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380:2197-223.
25. Harvard School of Public Health, World Health Organization, World Bank. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank; Distributed by Harvard University Press; 1996.
26. Finkelstein E, Corso P. Cost-of-illness analyses for policy making: a cautionary tale of use and misuse. Expert Rev Pharmacoecon Outcomes Res. 2003;3:367-9.
27. Colombia, Ministerio de la Protección Social. Guía metodológica para la elaboración de Guías de Atención Integral en el Sistema General de Seguridad Social en Salud Colombiano. Bogotá: Minprotección; 2010.
28. Rojas MX, Rodríguez VA, Tamayo DC, et al. Costos médicos directos del síndrome de dificultad respiratoria del recién nacido en Colombia. Coyunt Económica [internet]. 2013. [citado 2014 sep. 1]. Disponible en: http://ideas.repec.org/a/col/000438/012045.html
29. Costa N, Derumeaux H, Rapp T, et al. Methodological considerations in cost of illness studies on Alzheimer disease. Health Econ Rev. 2012;2:18.
30. Ripari NV, Moscoso NS, Elorza ME. Costos de enfermedades: Una revisión crítica de las metodologías de estimación. Rev Lect Econ [Internet]. 2012 [citado 2014 sep. 1]. Disponible en: http://ideas.repec.org/a/col/000174/010687.html

CAPÍTULO

24

Estudios de farmacoepidemiología y farmacoeconomía

Rodolfo José Dennis Verano

INTRODUCCIÓN

En este capítulo se tratan los temas de farmacoeconomía y farmacoepidemiología como estrategias complementarias para evaluar el impacto del uso de los medicamentos en seres humanos, y generar información que pueda ser útil en el proceso de toma de decisiones. Tanto los medicamentos como las vacunas son considerados tecnologías biomédicas, y los aspectos relacionados se incluyen en la sección de evaluación de tecnología en el sector salud.

No es evidente de inmediato la razón por la cual ambas áreas del conocimiento, tratadas en este capítulo, pueden ser complementarias, más allá de que ambas consideran *objeto* de estudio a los medicamentos. Al fin y al cabo, por un lado, se trata de los fundamentos de la epidemiología aplicados al estudio de los medicamentos[1], y, por otro, de los principios de economía de la salud aplicados al mismo objeto de estudio. Sin embargo, la evaluación económica del impacto de los medicamentos en el sector salud tiene como insumo fundamental, para las integraciones que lleva a cabo, la evaluación válida y confiable de la efectividad y del riesgo asociado a su uso en poblaciones humanas. Esto, regularmente, requiere datos que provengan de estudios no experimentales, ya cuando los medicamentos se hayan difundido a la comunidad tras las aprobaciones regulatorias para su utilización, y así se pueda evaluar mejor la efectividad, y no solo la eficacia comparativa.

Los diseños de tales estudios, sus fortalezas, sus debilidades y las medidas para la mejor prevención de sesgos son, precisamente, las áreas de interés de la epidemiología. Por otro lado, la economía de la salud y sus métodos han llevado más allá de la epidemiología la evaluación del impacto de los medicamentos en una sociedad, al integrar de manera completa los desenlaces intencionados, los desenlaces adversos, las consecuencias en costos asociados a ambos y el impacto global sobre la calidad de la vida, mediante la estimación de años de vida ajustados por calidad o por discapacidad, o en equivalentes de años de vida saludable.

EJEMPLO

En la evaluación completa del impacto del tratamiento de la infección por el virus de la hepatitis C[2] parece ser que hay nuevos medicamentos con verdadero potencial para erradicar tanto la infección como sus consecuencias (cirrosis, aparición de cáncer hepático, necesidad de trasplante, y muerte), probados en estudios experimentales iniciales, premercadeo o precomercialización. Sin embargo, si estos nuevos medicamentos, una vez utilizados en la práctica clínica usual, no mantienen las mismas proporciones de curación, o si se asocian a eventos adversos potencialmente severos (algunos solo detectables y cuantificables cuando un número mayor de sujetos se hayan expuesto después de su comercialización), su relación de beneficio-riesgo no sería del todo clara, comparada con el mejor estándar actual del manejo de la hepatitis C.

Un análisis de esta situación requiere, como insumo esencial, una estimación válida y confiable de la frecuencia de los desenlaces de interés, tanto intencionados como adversos, en situación de práctica clínica usual. Esta suele ser el área de interés de la farmacoepidemiología.

Por otro lado, para el nivel decisorio de un sistema de salud de pagar por el nuevo medicamento, esa información requiere ser más elaborada para tomar una decisión sustentada en la mejor evidencia. La decisión informada, probablemente, requiera saber, de forma comparativa contra el mejor estándar actual de tratamiento, cuánta es la efectividad incremental, cuántos los desenlaces adversos incrementales —provenientes de la farmacoepidemiología—, y cuántos, los costos diferenciales asociados a la adquisición de los medicamentos, los costos diferenciales asociados a los desenlaces adversos, y los gastos que el nuevo medicamento evitaría por menos cirrosis, cáncer y trasplantes hepáticos; todo esto, en comparación con el estándar actual. Dicho modelamiento del impacto de esta nueva tecnología es el área de interés de la farmacoeconomía.

En este capítulo, inicialmente se describirán de manera resumida el desarrollo de los medicamentos en seres humanos y el estado del conocimiento promedio sobre el beneficio y el riesgo con los que un nuevo medicamento cuenta al ser difundido en la comunidad. Posteriormente, se describirán los estudios disponibles en farmacoepidemiología para la evaluación de medicamentos. Finalmente, se describirán los diferentes estudios disponibles en la evaluación económica de los medicamentos.

La evaluación de los medicamentos es, esencialmente, de naturaleza multidisciplinaria, y también trasciende las competencias tanto de epidemiólogos como de economistas. En ocasiones, la visión limitada de ambas áreas del conocimiento puede llevar, en la práctica, a poca colaboración entre ambas. El objeto de estudio (la medición del impacto de los medicamentos en la sociedad) y la innovación en los métodos para llevarlo a cabo cada vez mejor requieren, claramente, una mayor sinergia entre la farmacoepidemiología y la farmacoeconomía[3].

EL PROCESO DE DESARROLLO Y APROBACIÓN DEL USO DE MEDICAMENTOS

El proceso de desarrollo de medicamentos en Europa y en Estados Unidos, típicamente, tiene una parte preclínica en animales y una parte clínica en seres humanos, la cual, a su vez, tiene tres fases. Por cada medicamento que aprueba la Food & Drug Administration (FDA) de Estados Unidos, aproximadamente 7000 compuestos han sido investigados inicialmente; solo 30 alcanzan a ser investigados para evaluar su toxicidad en animales, y 7 más llegan a ser experimentados en seres humanos[4]. El proceso demora, aproximadamente, 10 años, y puede costar más de 1000 millones de dólares, aunque la metodología para calcular dicho costo no es totalmente confiable[5].

En la fase de toxicología animal se busca aproximarse a la relación dosis-efecto, y evaluar la cinética; también, predecir el peligro y la toxicidad para el ser humano, definir posibles daños en los órganos blanco, y rechazar los compuestos inaceptables.

La fase clínica, de experimentación con seres humanos antes de la aprobación, usualmente, consta de tres fases **(tabla 24.1).**

Durante la fase I de experimentación en seres humanos, se expone a un número reducido de

Tabla 24.1 El proceso de desarrollo de medicamentos por la Federal Drug Administration (FDA) en Estados Unidos

Fase de desarrollo	Duración (años)	Número de medicamentos candidatos	Propósito principal	Número de sujetos humanos expuestos
Fase prehumana	6	Inicial: 7000 Estudios de toxicidad animal: 20	Dosis-efecto; cinética; predecir peligro y toxicidad; definir posible órgano blanco del daño	
Fase I en seres humanos	2	7	Toxicidad; cinética y metabolismo; rango inicial de dosis	10-50 sujetos
Fase II en seres humanos	2	3	Seguridad; definir dosis o rango de dosis efectiva para fase III	50-200
Fase III en seres humanos	3	2	Seguridad; eficacia en las diferentes indicaciones	500-3000
Fase IV posmercadeo	15	1	Seguridad; efectividad	Población blanco del tratamiento

Fuente: Ciociola, et al.[4]

ellos[10-50]; por lo general, sujetos jóvenes, sanos y voluntarios. Se trata de excluir la toxicidad obvia no predicha durante la fase preclínica, evaluar la farmacocinética y el metabolismo del medicamento, y determinar un rango inicial de dosis que sea seguro en seres humanos. En situaciones donde el medicamento sea citotóxico, esta fase se hace directamente en los pacientes, como en el caso de medicamentos para el cáncer.

La fase II se hace directamente con sujetos que tienen la enfermedad de interés. El principal énfasis es definir una dosis efectiva que brinde el mejor perfil de riesgo-beneficio para los experimentos de fase III, y excluir los efectos adversos más comunes. Por lo general, se expone al medicamento a, aproximadamente, 200 sujetos.

Durante la fase III se hacen estudios aleatorizados de un número mucho mayor de sujetos (aproximadamente, 2000-3000), en quienes se trata de confirmar los resultados de seguridad y eficacia vistos a lo largo de la fase II, en las diferentes indicaciones del medicamento. Aún si la eficacia se puede demostrar en un menor número de pacientes en esta fase, se trata de tener suficiente precisión para detectar y cuantificar eventos adversos que se presenten con frecuencias > 1 en 1000[6].

Una vez terminadas estas fases, el patrocinador recopila toda la información que crea pertinente sobre manufactura y estudios preclínicos y clínicos, y la somete a las agencias gubernamentales pertinentes. Al final de este proceso, y cuando el medicamento ha sido evaluado y aprobado por la agencia responsable, el medicamento se puede comercializar. Si bien se espera que en un futuro entidades como la Agencia de Alimentos y Medicamentos (FDA, por las siglas en inglés de Food and Drugs Administration) y la Agencia Europea de Medicamentos (EMA, por las siglas en inglés de European Medicines Agency) exijan controles posmercadeo más estrictos[7], hay limitaciones actuales de importancia respecto al estado del conocimiento con el que los medicamentos difunden al mercado. Estas incluyen:

- Su efecto en los pacientes con comorbilidades importantes activas ha sido poco evaluado.
- La situación terapéutica en la práctica puede diferir, por diferentes grados de severidad.
- Su efecto en otros perfiles de pacientes como ancianos, niños y mujeres embarazadas es poco conocido.
- Su afecto a largo plazo es poco conocido.
- Su impacto con la administración concomitante de otros medicamentos ha sido poco evaluado.
- Sus eventos adversos poco comunes, pero potencialmente severos, no se conocen o no se han cuantificado.

La tabla 24.2 trae el conocimiento promedio sobre diferentes atributos de importancia que se tiene sobre un medicamento en el momento

Tabla 24.2 Estado promedio del conocimiento de diferentes aspectos respecto a un medicamento en el momento de difusión al mercado

Parámetro de interés	Estado del conocimiento
Seguridad	• **Preclínica: bastante bueno** • Clínica: bueno • Perfil de eventos adversos frecuentes: bueno
Farmacología clínica	Regular
Indicaciones	Regular; pueden cambiar con la experiencia
Subgrupos de mayor o menor beneficio	Regular
Dosis	Regular; ocasionalmente, incorrecta
Método global de uso	Regular; a menudo, pobremente desarrollado
Efectividad	Regular
Costo-efectividad	Por lo general, desconocida
Cuantificación de riesgo	Por lo general, no establecida

de ser registrado y aprobado. Pese al estado del conocimiento regular que se tiene, es claro que en vez de demorar la difusión de nuevos medicamentos (después de 10 años, en promedio), quizás, la mejor opción es dejar difundir los medicamentos al mercado. Esto, claramente, va de la mano con tener una fase IV posmercadeo para mejorar la seguridad asociada a la utilización de los medicamentos, mediante estudios que, en su mayoría, no son experimentales, y que alerten de forma temprana sobre eventos no intencionados o deletéreos para el ser humano. En esencia, este es el campo de acción de la farmacovigilancia y de la farmacoepidemiología.

FARMACOEPIDEMIOLOGÍA

La farmacoepidemiología (FE) es el estudio de los medicamentos como determinantes de salud y enfermedad en poblaciones humanas[6]. El *foco* de su estudio cae dentro del ámbito de la farmacología clínica, y sus *métodos* son los de la epidemiología[8].

Su mayor interés recae en ***eventos*** o desenlaces asociados a la utilización de medicamentos en grupos poblacionales, y no simplemente en individuos. Un evento es cualquier efecto, deseado o indeseado, asociado al uso de un medicamento, y que no es el efecto farmacológico intencionado. Un evento ***adverso*** es cualquier experiencia médica desfavorable que pueda aparecer durante el tratamiento con un medicamento, sin que tenga ***necesariamente*** relación causal con dicho tratamiento (por ejemplo, contraer un estado gripal, sufrir una fractura durante un partido de fútbol, tener un episodio de cálculos renales…). El fundamento de la definición estriba en que habitualmente no se sospecha sobre ese evento adverso, no se tiene información sobre relación causal con el tratamiento y esa relación causal no tiene que estar demostrada.

Por otro lado, por farmacovigilancia (FV) se entiende el conjunto de actividades destinadas a identificar ***reacciones*** adversas previamente desconocidas, cuantificar su riesgo causal, tomar medidas reguladoras al respecto e informar a los profesionales de la salud y al público en general[9].

Una reacción adversa se define como un efecto deletéreo no intencionado que ocurre cuando un medicamento, administrado a la dosis usual recomendada, se utiliza con el propósito de prevenir, diagnosticar o tratar. En este caso, el efecto puede predecirse a menudo (sabor metálico en la boca con el metronidazol, gastritis con ciertos antibióticos, dolores musculares con las estatinas…), y para llamarlo ***reacción adversa***, se debe tener evidencia de la relación causal.

Adicionalmente al foco en reacciones adversas, y menos en eventos adversos, la inferencia causal y los métodos utilizados también ayudan a distinguir a los campos de la FV y de la FE. En la FV, la evidencia recogida sobre reacciones adversas tiende a ser determinística, se preocupa por la imputabilidad de la reacción al medica-

mento (posible, probable, definitiva) y su inferencia se hace desde el individuo afectado; en la FE, **la evidencia recogida es probabilística, se** preocupa por cuantificar la extensión del riesgo y su inferencia es desde grupos poblacionales expuestos y no expuestos.

La metodología más relevante en la FV es la notificación activa o pasiva de potenciales reacciones adversas; las notificaciones generan señales, numeradores, de los cuales se desprenden hipótesis de investigación. Los pasos siguientes, tales como estimar el riesgo de ocurrencia de eventos adversos, refutar hipótesis e identificar grupos de mayor o de menor riesgo, son, usualmente, el campo de la FE.

Sesgos en FE

En esencia, el control de sesgos es la principal contribución de la epidemiología al área de estudio del riesgo asociado a la utilización de medicamentos en grupos poblacionales[10].

Los sesgos operantes en los estudios de FE son los mismos que en otros estudios epidemiológicos, pero con situaciones particulares.

Sesgos de ***selección***, **que operan al seleccionar** a sujetos para los estudios, o al perderlos durante el seguimiento en el tiempo; sesgos de *información*, **o de mala clasificación, que actúan** al medir el uso del medicamento o el evento de interés, o las posibles variables cointervinientes, y sesgos de ***confusión***, en los que un factor extrínseco a la cadena causal entre el medicamento y el desenlace de interés explica, en parte o totalmente, la asociación (o la ausencia de ella).

Sesgos de selección

En este tipo de sesgo, hay una distorsión del estimativo del efecto, debido al ingreso al estudio de sujetos con una relación desigual e inusual entre el uso del medicamento y el desenlace de interés. La muestra no es, entonces, representativa de la población blanco, y el estimativo se invalida.

Existen otros ejemplos de sesgos de selección en estudios de FE:

- Por referencia: **El uso del medicamento facilita** que el paciente sea llevado a algún examen diagnóstico, así como referir al paciente para más estudios o para hospitalización. En tal estado, la presencia de la exposición al medicamento hace más probable, entonces, que se incluya a estos pacientes en la investigación. Como ejemplos de ello están referir para endoscopia, preferencialmente, a personas con exposición a antiinflamatorios, u hospitalizar para estudio de posible trombosis venosa, preferencialmente, a mujeres con exposición conocida a los anticonceptivos orales.
- Por autoselección: **El sujeto investigado decide** ingresar (o dejar) el estudio sin filtros; la decisión de hacerlo puede estar relacionada, simultáneamente, con la exposición al medicamento y a desarrollar el desenlace de interés. Como ejemplos de ello, las madres de hijos con defectos de nacimiento suelen declararse más dispuestas a ingresar a un estudio de posibles medicamentos asociados; así mismo, la participación voluntaria de pacientes con glaucoma o cataratas se verá estimulada si han hecho uso previo de esteroides inhalados.
- Por casos prevalentes: **En estudios epidemiológicos** de casos y controles, la inclusión de personas supervivientes (prevalentes) a la enfermedad bajo estudio puede no reflejar la magnitud de la asociación a un medicamento; si los casos más severos, potencialmente mortales, son los asociados al medicamento, no estarán bien representados en una muestra de sujetos con tiempo largo desde el momento del diagnóstico. Un ejemplo podría incluir el de anorexígenos orales y el desarrollo de hipertensión pulmonar severa.
- Protopático: **Una decisión específica (dar un** medicamento) es empezada, suprimida, o modificada por la manifestación inicial,

EJEMPLO

En un estudio clínico, los sujetos son reclutados gracias a la referencia que hacen médicos de consulta ambulatoria desde consultorios de diversas zonas de la ciudad. Si los médicos han sido sensibilizados previamente (por medio de una carta de distribución general) respecto a la posible asociación entre el uso de un medicamento en estudio y un evento adverso, será mayor la probabilidad de que se los detecte, y ello, claramente, influirá en cuáles pacientes serán remitidos para el estudio.

usualmente sutil, del desenlace bajo estudio, lo que lleva a detectar asociación, ***pero casual***. **Como ejemplos, hay estudios de** casos y controles acerca del inicio de estrógenos por sangrado uterino, cuando esta es la incipiente manifestación de un carcinoma de endometrio no diagnosticado; o la iniciación de aspirina que posteriormente se asocie a sangrado rectal y, eventualmente, lleve al diagnóstico de carcinoma de colon.

Sesgos de información (mala clasificación)

Es raro, por no decir imposible, tener un instrumento que clasifique correctamente al 100 % de las exposiciones a medicamentos o de los desenlaces de interés. La gravedad del sesgo es función de la dependencia —o no— entre la mala clasificación de exposición y el evento de interés[11]. **Estimar de forma inadecuada la** *función de riesgo* **es una forma especial de este** sesgo; asumir como ventana de exposición periodos muy largos, fuera de la verdadera ventana causal, lleva a la dilución del efecto, e, incluso, a perder asociaciones reales. Un ejemplo sería el de la asociación entre exposición a antiinflamatorios no esteroideos (**AINE**) **y hemorragia de las vías** digestivas: utilizar exposición a **AINE en los seis** meses anteriores como relacionada, posiblemente, con la hemorragia de vías digestivas.

Sesgos de confusión

En estos sesgos, por desgracia comunes, el estimativo del efecto entre la exposición al medicamento y el desenlace se ve distorsionado por el efecto de una o más variables que también se asocian, o son pronósticas, del desenlace. El sesgo de confusión surge cuando dicho factor no está balanceado, o repartido por igual, entre las categorías de exposición. Un ejemplo de confusión por edad sería asociar la exposición a la teofilina y el desarrollo de arritmias supraventriculares. A mayor edad hay un mayor desarrollo de estos trastornos del ritmo. A mayor edad hay mayor probabilidad de formulación de teofilina. No ajustar en el diseño o en el análisis por el efecto de la edad podría llevar a que se sobreestime el riesgo entre teofilinas y el desarrollo de arritmias.

Otra posibilidad, potencialmente común, es el sesgo por la ***contraindicación*** **hacia el** medicamento en estudios observacionales (no experimentales); usualmente, casos y controles. En este caso, la historia médica de una condición, o de un síntoma, disminuye la formulación del medicamento por los prescriptores en pacientes percibidos como más vulnerables al riesgo de desarrollar el desenlace de interés. El sesgo puede ser tan marcado que, en ocasiones, la asociación observada tiende hacia la disminución del riesgo, cuando el efecto real es lo contrario. Un ejemplo de ello es cuando se evalúa la asociación entre **AINE y el sangrado gastrointestinal** en estudios de casos y controles. Si no se excluye del estudio a los ya conocidos susceptibles a su efecto (variable de confusión), sería mucho menos común la frecuencia de exposición a **AINE** en pacientes con, digamos, dispepsia previa, y que han presentado el desenlace de sangrado. La frecuencia de exposición global a **AINE** puede, incluso, ser mayor en el grupo control sin sangrado.



Finalmente, existen sesgos de confusión asociados a la ***indicación*** **por el medicamento, al evaluar los** efectos ***intencionados*** de un medicamento a través de diseños no experimentales, como estudios de cohortes o estudios de casos y controles.

La indicación médica de uso de muchos medicamentos es, a la vez, factor pronóstico del desenlace adverso que se está tratando de prevenir con el uso del medicamento, aun si el medicamento es efectivo. En ausencia de aleatorización, la indicación ***per se*** **puede constituirse** en variable de confusión: está —obviamente— relacionada con la prescripción del medicamento, y, simultáneamente, es un factor pronóstico de la ocurrencia del desenlace de interés. Un ejemplo es la indicación del uso de heparinas para prevenir la trombosis venosa en pacientes hospitalarios de cirugía ortopédica mayor. Cuanta mayor sea la percepción de riesgo en ausencia de profilaxis, mayor es la probabilidad de prescribirla, pero, al mismo tiempo, también será la ocurrencia de eventos en pacientes de muy alto riesgo, incluso en presencia de profilaxis efectiva, cuando se compara con pacientes que no recibieron profilaxis. Esto puede llevar a atenuar la medición del verdadero impacto benéfico de la profilaxis.

Obviamente, en presencia de estudios experimentales aleatorizados, la confusión por la indicación estaría controlada, pero en los

estudios no experimentales es uniformemente difícil ajustar por el efecto de la indicación[12], y a menudo queda la duda de confusión residual por la indicación que llevó a la prescripción.

Conceptos básicos de diseños en FE para evaluar el riesgo

El objetivo principal de la FE es la cuantificación posmercadeo del riesgo asociado al uso de medicamentos, y la inferencia causal entre la exposición al medicamento y la ocurrencia del evento adverso[13]. En tales condiciones, la aproximación experimental es muy poco factible, y se debe recurrir a diseños observacionales.

Sin embargo, antes de intentar diseñar el estudio no experimental, es útil tener claros los siguientes puntos:

- La aproximación no experimental es factible: **Hay** baja probabilidad de que las manifestaciones iniciales, y sutiles, del desenlace no hayan llevado a iniciar o a modificar la exposición al medicamento.
- La duración de la exposición relevante para el desenlace: **Usualmente, incide en la inducción del** evento adverso y en el exceso de riesgo.
- La probabilidad de exposición previa al medicamento: **Un evento adverso anterior, por lejano** que haya sido en el tiempo, constituye en medicina clínica una contraindicación para la reexposición.
- La consideración de dos periodos componentes: **El** riesgo de la ocurrencia del desenlace durante una exposición continuada al medicamento, y el riesgo después de haberse descontinuado. Cuando el tiempo de exposición es largo respecto al tiempo de inducción del evento adverso, solo es pertinente cuantificar el tiempo de exposición; por otro lado, cuando el tiempo de exposición al medicamento es corto, virtualmente todo el riesgo estará dado por el periodo postexposición.
- La ventana de riesgo que se estima como plausible: **Es poco probable que el riesgo asociado** a la ocurrencia del evento adverso se mantenga constante en el tiempo. Por ejemplo, la función de riesgo durante el periodo de exposición puede aumentar inicialmente, y luego disminuir por depleción de susceptibles. Durante el periodo de descontinuación, el riesgo, eventualmente, puede alcanzar el riesgo referente, y el exceso de riesgo, en esencia, ser cercano a cero. Por ende, el riesgo tras descontinuar el medicamento estaría limitado a un periodo singular, que debe ser diseñado de manera que cubra la mayoría de los eventos atribuibles a la exposición reciente. Desafortunadamente, tal periodo se desconoce, y en el análisis deben evaluarse varios segmentos de un periodo ***a priori*** excesivamente largo.

En resumen, independientemente del diseño, delimitar el problema implica cuantificar las densidades de incidencia del evento adverso en discusión, para varios segmentos de tiempo durante la exposición y después de ella; también, cuantificar las densidades de incidencia del evento adverso en ausencia de exposición reciente, con particular atención a las características del grupo expuesto, y cuantificar, mediante una medida de asociación (como riesgo relativo o diferencia absoluta de riesgos), si no hay sesgo protopático (o si se controla en el diseño) y si no hay confusión por la contraindicación (es decir, si no hay historia del evento en exposiciones previas).

Diseños de investigación en FE

Dado que los diseños de investigación en **FE** son, en esencia, los mismos de la epidemiología clínica, estos se revisan en detalle en el **capítulo 13** "Introducción a los diseños en investigación: selección del diseño apropiado", de este libro. En esta sección, se revisarán, de forma somera, los diferentes diseños analíticos y descriptivos de uso en **FE** **(tabla 24.3)** asumiendo que el lector ya ha consultado los capítulos respectivos sobre diseños, en este libro. Adicionalmente, se le recomienda al lector la consulta de libros de texto[14].

En FE, **el *experimento clínico aleatorizado*** (**ECA**), al igual que en otras áreas del conocimiento, es el paradigma actual de la investigación sobre efectos intencionados en la salud o la enfermedad. La presencia de aleatorización controla sesgos al ensamblar los grupos, permite asumir comparabilidad en variables extrínsecas a la cadena causal y evita la confusión por la indicación del tratamiento. Sin embargo, en la evaluación de desenlaces no intencionados (en especial, los de eventos adversos con el uso

Tabla 24.3 Tipos de diseños de estudios en FE

Estudios analíticos
• Experimento clínico controlado
• Cohortes: concurrentes, históricas, o ambidireccionales
• Casos y controles de base primaria o secundaria
• Caso-cohorte
• Caso-cruzado: caso-tiempo-control, caso-caso-tiempo, serie de casos con autocontrol
Estudios descriptivos
• Reportes de caso
• Series de casos
• Ecológicos
• Corte transversal

de medicamentos) la exposición intencionada de los sujetos de investigación a medicamentos no es, por lo general, éticamente viable; así mismo, puede no ser eficiente, dada la rareza de la mayoría de los desenlaces evaluados en FE.

Si asumimos que el ECA no es ético o no es viable, una opción es investigar una *cohorte* de personas expuestas; cada una, "matriculada" en la cohorte con el comienzo de la exposición al medicamento, y seguidas durante el periodo definido como de riesgo (*ventana de interés*). Esta cohorte de pacientes expuestos sería suplementada por una cohorte de personas no expuestas, y seguidas a lo largo de un periodo similar. Tal acercamiento al problema también puede ser bastante impráctico: la rareza de los eventos adversos requieren el ensamblaje *de novo* de cohortes gigantescas, y en casos de exposiciones durante periodos largos, o periodos de inducción y de latencia largos, se necesitaría seguir la cohorte por largo tiempo, con los costos y el potencial de pérdidas que ello supone.

Los mencionados problemas pueden atenuarse si se utilizan cohortes cerradas o poblaciones dinámicas ensambladas para otros propósitos[15]. En esta situación, pueden definirse y ensamblarse cohortes en las que se sigue la experiencia en persona-tiempo de los sujetos expuestos y no expuestos a través de bases de datos computarizadas ensambladas primariamente para fines administrativos[16]. Usar estas bases de datos administrativas ha revolucionado la práctica de la FE para la evaluación de riesgo[17]. Existen, sin embargo, algunas precauciones generales que se deben llevar a cabo antes de su utilización (hay de por medio componentes críticos mínimos, necesarios antes de considerar su uso)[16]:

1. Disponibilidad de ítems críticos, tales como la fecha del servicio prestado, el procedimiento realizado, los detalles de la prescripción y los datos demográficos generales.
2. Validación independiente de estos datos, a través de una revisión directa de la historia clínica, o del tercero pagador, o a través del paciente mismo (esto último, poco común).
3. Factibilidad:
   - Si el número de sujetos en la base administrativa es < 20 000 en sujetos mayores de 60 años, es poco probable que se pueda discriminar entre resultados verdaderamente negativos de falsos negativos (*poder estadístico bajo*) para muchos eventos adversos de rara ocurrencia.
   - Representatividad respecto a la población blanco del investigador.
   - Estabilidad: ¿cuánto tiempo lleva funcionando?
   - Integridad: datos de diagnósticos en códigos de la International Classification of Diseases (ICD); datos de procedimientos codificados; medicamentos codificados; código de trabajador de la salud.
   - Disponibilidad para investigación.

Lo poco común del evento adverso en estudios de FE, si bien ha mejorado mucho su análisis con bases de datos administrativas computarizadas que permiten ensamblar grandes cohortes sin contactar directamente a la persona, pueden ser todavía ineficientes; en especial, cuando se debe validar externamente parte de la información.



Para mejorar la eficiencia del estudio de eventos adversos poco comunes en FE, se han desarrollado diseños que utilizan *muestras* de las cohortes. Los más conocidos son el estudio de *casos y controles anidado* (o de base primaria), y el estudio de *caso-cohorte*.

El diseño de casos y controles anidado en cohortes[18] involucra cinco pasos principales:

- Definir el *eje de tiempo de la cohorte*: tiempo calendario o tiempo de seguimiento (inusual).
- Seleccionar los casos del evento en la cohorte: todos (más usado) o una muestra aleatoria

(menos usado; usualmente, por el limitado número de casos disponible).
- Formar los grupos de riesgo correspondientes: incluye los niveles de ***expuesto*** y ***no expuesto***, así como los criterios de pertenencia a la cohorte.
- Escoger al azar la muestra de uno o más controles en los grupos de riesgo escogido: incluye los criterios de emparejamiento que se utilizarán: por ejemplo, puntajes de propensión[19] y el tipo de muestreo (por ejemplo, por densidad de incidencia).
- Análisis de datos condicionado por el tipo de muestreo y emparejamiento utilizado.

El estudio de caso-cohorte se puede percibir como una minicohorte con todos los casos agregados, o un estudio de casos y controles anidado, pero sin emparejamiento, e involucra, a su vez, tres pasos principales:
- Seleccionar todos los casos de ocurrencia del evento.
- Escoger al azar un número precalculado de sujetos como grupo de contraste.
- Un análisis estadístico condicionado por el tipo de diseño.

Como alternativa a diseñar estudios de casos y controles anidados en bases primarias, el investigador se puede aproximar a la experiencia de persona-tiempo de los sujetos expuestos al medicamento, a través de la forma como son identificados los casos del desenlace (por ejemplo, a través del área de influencia de un hospital o de un grupo de hospitales).

En tal situación, el estudio de casos y controles se considera de ***base secundaria,*** ya que los casos pueden provenir de bases de datos hospitalarias computarizadas[20]. En este tipo de diseño, la identificación de los casos se puede hacer de forma sistemática, mediante la búsqueda del desenlace en la historia clínica electrónica (evento adverso) en los diagnósticos de egreso hospitalario; a través de los archivos electrónicos de farmacia, se identifica a los sujetos expuestos al medicamento, y el cruce entre la información obtenida de la historia clínica y la del archivo de la farmacia sirve para determinar el estatus de exposición en los casos. Por otro lado, la identificación de los controles se haría también a partir de los egresos hospitalarios (como representantes de la base secundaria), y se seleccionaría una muestra al azar emparejada (edad, sexo, fecha de ingreso, puntajes de propensión) con los casos; a través de los archivos electrónicos de farmacia se identifica a los sujetos expuestos al medicamento, y el cruce entre la información de egresos hospitalarios y la obtenida del archivo de farmacia sirve para determinar el estatus de exposición en los controles. Luego, el análisis estadístico de los datos de ambos grupos procedería como descrito para ese tipo de estudios.

Para el estudio de efectos medicamentosos *transitorios* que pueden generar eventos adversos agudos en el tiempo, es posible que la mejor representatividad de la población blanco del investigador, como grupo de referencia o de contraste, sea el propio paciente.

(Por ejemplo, el estudio de arritmias con la exposición a la inhalación de broncodilatadores agonistas de los receptores beta2). Para tal tipo de situaciones, solo los estudios en FE con casos pueden ser de utilidad[21]. El diseño más habitual se conoce como ***caso-cruzado***[22,23]. El diseño y sus variantes (caso-tiempo-control, caso-caso-tiempo, serie de casos con autocontrol) son conceptualmente similares al experimento clínico cruzado. Este tipo de diseño eliminaría sesgos de selección y de confusión en el mismo sujeto por factores que no varíen de forma importante en el tiempo, pero solo es útil en eventos adversos agudos como resultado de efectos transitorios, con comienzo abrupto y claramente identificable. Requiere, además, conocimiento preciso de la ventana de acción del efecto para poder asignar de forma creíble el estatus de expuesto/no expuesto. El diseño de dichos estudios involucra:
- Identificar en la cohorte todos los casos del desenlace de interés.
- Determinar la ventana probable del efecto basado en la farmacocinética del medicamento (***riesgo de desarrollo del evento***).
- Obtener el patrón de uso del medicamento para cada caso identificado, por un tiempo que se considere suficiente.
- Determinar, para cada caso, si el evento ocurrió durante la ventana de exposición, o por fuera de ella.
- Como contraste, identificar cuántos periodos de exposición y de no exposición pudo haber

habido en el tiempo definido ***a priori*** como suficiente (por ejemplo, un año).
- Generar, para cada caso, una tabla de 2 × 2 con la información recopilada, y analizarlo como se haría para un análisis estadístico estratificado.

No es un secreto que el sector del cuidado de la salud está en constante evolución. Hay muchos determinantes de dicha evolución, que incluyen, entre otros, las nuevas tecnologías; especialmente, medicamentos y dispositivos[24]. Se espera del sector de la salud, por parte de la sociedad, que haya disponibilidad casi inmediata de muchas tecnologías que mejoran o prolongan la vida, y que se provea acceso amplio e irrestricto al sistema a la mayor cantidad posible de usuarios, pero, a la vez, que se contengan los costos del sistema dentro de ciertos límites. En una situación de recursos finitos, esas tres responsabilidades del sector de la salud en un gobierno están claramente en conflicto.

Surge, entonces, la pregunta sobre cuál es la mejor forma de asignar equitativa y éticamente los recursos existentes en el sector, para lograr el máximo beneficio de la sociedad. Esta es el área de estudio y de aplicación de la economía de la salud: ayudar en la toma de decisiones basadas en la mejor evidencia[25]. Una de las áreas de aplicación de la economía de la salud es el área interesada en el análisis económico del impacto de los medicamentos, o la ***farmacoeconomía***

La farmacoeconomía (FC) puede definirse como la aplicación del análisis económico al campo de los medicamentos. Es la determinación de la eficiencia (relación entre costo y efecto) de un tratamiento medicamentoso y su comparación con las otras opciones, con el fin de seleccionar las de una relación costo-efecto más favorable.

Esta área del conocimiento combina las disciplinas de economía de la salud, epidemiología, análisis de decisiones, farmacología clínica y bioestadística[26]. Es claro que todos los entes involucrados en el ámbito decisorio sobre medicamentos tienen la capacidad de beneficiarse del análisis económico minucioso de medicamentos que están difundiendo en el mercado: las agencias reguladoras, el sistema de salud, los aseguradores, la industria farmacéutica y la sociedad en general[27].

Es claro, además, que comparar el precio cobrado por un medicamento con el precio de otro medicamento que también esté indicado en la misma situación no abarca todo lo que se necesita saber sobre el impacto económico que tiene un medicamento respecto al otro.

Este modo simplista de ver el impacto económico deja por fuera fuentes de costos tales como:
- Ahorro para el sistema, por menos desenlaces adversos, si uno es más efectivo que el otro.
- Ahorro para la sociedad, por tener a personas más saludables.
- Costos asociados al seguimiento del impacto del medicamento.
- Costos asociados a monitorizar la aparición de eventos adversos.
- Costos asociados al tratamiento de eventos adversos.
- Precio ***per se*** que se está dispuesto a pagar por estar más saludable.

De lo anterior se desprende que el mayor costo de adquisición de un nuevo medicamento puede terminar en ganancia neta para la sociedad si ahorra recursos por tener mayor efectividad o menor perfil de eventos adversos que el medicamento comparador.

Cabe anotar que si bien los economistas de las salud pueden ayudar a diseñar los estudios, son los consumidores de estos (como el gobierno, las entidades promotoras de salud —EPS—, las sociedades científicas) los que deben tomar las decisiones, y aceptar que gastar recursos en una situación o una enfermedad evita que tales recursos estén disponibles para su siguiente mejor uso ("costo de oportunidad" de la intervención). De ahí la necesidad de tener una visión más amplia que la simple contención de costos a cualquier precio (incluyendo no tener en cuenta la efectividad de estas intervenciones).

Cuándo hacer el análisis económico de medicamentos

Este punto tiene dos componentes; el primero es cuándo estaría indicado hacer un análisis

económico a un medicamento, y la segunda es si, estando indicado, cuándo en el ciclo de vida de un medicamento se debe hacer dicho análisis. Respecto al primer punto, es imposible hacerle o exigirle una evaluación económica completa a todas las intervenciones en salud candidatas a ser implementadas en la práctica clínica, incluyendo los medicamentos. Para priorizarlas, por tanto, se sugieren algunas consideraciones[28]:

- Las alternativas de medicamentos enfrentados afectan desenlaces relevantes (por ejemplo, mortalidad o desenlaces cardiovasculares mayores).
- Hay suficiente evidencia de calidad elevada respecto a la eficacia y la efectividad como para informar el proceso de análisis económico.
- Las evaluaciones preliminares sugieren que puede haber diferencias incrementales considerables en costos y adherencia entre las alternativas enfrentadas.
- No hay evidencia clara ***a priori*** de la dominancia de una alternativa (pero podría establecerse luego de una evaluación económica).
- Hay una elevada variabilidad en la práctica actual en la escogencia de las alternativas enfrentadas.
- El cambio en la práctica podría traer marcados beneficios en salud.

El cambio en la práctica podría racionalizar sustancialmente la utilización de recursos y contener costos innecesarios.

Respecto al segundo punto, y asumiendo una situación con un nuevo medicamento no comercializado aún, pero avanzado en la fase de desarrollo (fase II o III en seres humanos), surge la pregunta de cuándo hacer análisis económicos de medicamentos: si antes o después de la difusión al mercado. Existen argumentos a favor y en contra de ambas opciones.

A favor de hacerlo antes de la difusión, desde el punto de vista social, no ha existido mayor consumo de recursos; además, no hay grupos políticos o de interés ya posicionados; se tiene, o se va a tener, la mejor evidencia de eficacia disponible; la medición del consumo de recursos es, por lo general, menos complicada si se hace conjuntamente con un experimento clínico pragmático en fase III, y los costos globales de hacer el estudio, probablemente, sean menores. Adicionalmente, cada vez más las agencias regulatorias están viendo de manera más favorable para aprobación medicamentos con un perfil favorable de eficiencia cuando son presentados.

En contra de hacerlo antes de la aprobación regulatoria está el hecho de que se necesitan modificaciones y mayor logística, lo cual complica la realización de experimentos clínicos; además, hay carga adicional para los investigadores, y puede haber muchos supuestos que necesiten confirmación (incluso, de las poblaciones que se exponen al medicamento antes y después de la difusión), y, por ende, el perfil de costo-efectividad del medicamento, una vez difundo a la comunidad, puede no ser el mismo; es decir, puede no ser válido ni preciso.

A favor de hacer análisis económicos tras la aprobación y la difusión del medicamento, cabe observar que la situación de ***efectividad*** (más que la de ***eficacia***) está asegurada: es, casi, la única manera confiable de aproximarse realmente al impacto del medicamento en la población observada de distribución de este (y no la esperada), y es, virtualmente, la única manera de identificar, evaluar y cuantificar los costos secundarios asociados a eventos adversos poco comunes, como insumo en el análisis del impacto económico real del medicamento bajo evaluación.

En contra de hacerlo después de la aprobación están los sesgos asociados a la estimación de efectividad derivada de estudios no experimentales en la fase IV, tales como la confusión por la indicación[10], que la medición válida y confiable del consumo real de recursos asociados al uso del medicamento es mucho más difícil, y, finalmente, que el hecho de retirar, restringir o limitar el uso del medicamento por baja "eficiencia" tras la aprobación y la difusión en el mercado es uniformemente difícil.

Las alternativas para la toma de decisiones respecto al momento de elaborar los estudios farmacoeconómicos en el ciclo de vida del medicamento podría basarse en verlo como un continuo, con diferentes objetivos y teniendo en cuenta las limitaciones inherentes.

Para medicamentos de los cuales se estima que serán difusores rápidos, en fase III temprana es factible un análisis económico preliminar; los insumos importantes en esta fase incluyen el tener clara una variable de efectividad que sea creíble, identificar las variables clave necesarias para cuantificar el consumo de recursos y modelar extensamente, y complementar con análisis de sensibilidad los puntos de mayor incertidumbre.

Al final de la fase III, con un experimento clínico pragmático especialmente diseñado para tomar en cuenta el análisis de costos y desenlaces, es más factible un análisis económico completo que se aproxime a la ocurrencia futura de costos y desenlaces en fase IV.

El contraste sería contra el estándar actual del manejo de la enfermedad, y el tamaño de la muestra tendría en cuenta la necesidad de medir la calidad de vida y el consumo de recursos. En fase IV, si se requiere, el modelo económico se puede complementar con nueva información (especialmente, de efectividad "comunitaria" en poblaciones especiales, tales como extremos de la vida, y poblaciones de diferentes tipos de aseguramiento del cuidado). Una combinación de datos experimentales con datos posmercadeo pueden unirse en el modelo final.

Principios y tipos de estudios en FC

Los principios del análisis económico en salud se hallan, en general, bien establecidos. Hay publicaciones y libros recientes disponibles que explican el detalle de dichos principios[29]. En esta sección brindaremos un resumen de ellos, aplicados a la FC. De igual forma, los capítulos 23 "Introducción a la evaluación económica de tecnologías en salud" y 29 "Plan de análisis", en este mismo libro, ayudan a complementar el tema.

En un análisis económico típico sobre medicamentos, los insumos necesarios asociados al uso de un medicamento se cuantifican (esto incluye el costo de adquisición del medicamento, así como los insumos y el recurso humano necesarios para administrarlo, si ello es pertinente), al igual que los desenlaces obtenidos asociados a su utilización (tanto benéficos como adversos), y se comparan contra alguna otra manera también aceptable de brindar el tratamiento. Si la ganancia incremental en desenlaces justifica el costo incremental, o si es menos costoso brindar el tratamiento, se considera satisfactorio, o, incluso, dominante, desde el punto de vista del análisis económico.



Por ende, un análisis económico completo requiere la *comparación* contra una manera diferente de brindar la atención que se está evaluando; una medida aislada de los costos y las consecuencias de una intervención significa poco, y solo es relevante si se la compara de forma explícita contra una manera alternativa de brindar la misma atención.

Por ejemplo, el costo y la efectividad de la hidroclorotiazida para el manejo de la hipertensión esencial solo es importante si se compara contra los resultados en costo y en efectividad obtenidos con, digamos, inhibidores de la enzima convertidora de angiotensina (IECA), bloqueadores del receptor de la angiotensina (BRA II), anticálcicos y bloqueadores beta.

Adicionalmente, la *perspectiva*, o punto de vista, del análisis económico debe ser claro. Esto, a su vez, ayuda a delimitar el tipo de costos que se tomarán en cuenta; lo que puede verse como costos desde un punto de vista se puede verse como *ahorro* desde otro. Por ejemplo, los antipsicóticos modernos permiten, en muchas ocasiones, el manejo ambulatorio efectivo y seguro del paciente esquizofrénico crónico institucionalizado; desde un punto de vista del asegurador, el medicamento sería costo-efectivo, al no obligarlo a incurrir en los costos de la atención institucionalizada; pero desde el punto de vista del análisis por el paciente o su familia, quizás no sería lo más eficiente si ello se asocia a una convalecencia larga en casa.

En esencia, hay cuatro grandes estrategias que pueden ser utilizadas para la evaluación económica de una tecnología en el sector de la salud.

En los estudios de *minimización de costos*, el consumo de recursos asociados a brindar un tratamiento se comparan contra el consumo de recurso asociados a utilizar un comparador. Aunque hace algunos años muchos análisis utilizaban ese diseño, es limitado e incompleto, pues el análisis no provee una medida creíble de los desenlaces asociados a los dos tratamientos. Como tal, el evaluador debe justificar por qué considera que los resultados son equiparables, y, por ende, que es irrelevante su cuantificación formal. Algunos autores no consideran que tal tipo de análisis sea una evaluación económica completa, dada, precisamente, la ausencia simultánea de comparación de los desenlaces obtenidos.

En el análisis de *costo-beneficio*, el consumo de recursos relevantes asociados a la intervención se identifica, se mide y se le asigna un valor, similar al tipo de estudio anterior. En este diseño, a los desenlaces también se les asigna un valor monetario. El resultado es, entonces, una razón de costos a beneficios, o un valor neto de la operación, dependiendo de si la intervención ahorra o no plata. Esta comparación evalúa,

EJEMPLO

Un estudio comparó una política hospitalaria de restringir el uso de algunos antibióticos contra su libre formulación, en un estudio de antes y después[30]. Los autores documentaron un ahorro neto por el uso de un aminoglucósido menos costoso, menor uso general de cefalosporinas de generación avanzada, y menor uso de vancomicina, de hasta 200 000 dólares mensuales. Los autores manifiestan en la discusión que ni el promedio de estancia ni la mortalidad hospitalaria variaron entre los dos periodos. Sin embargo, la evidencia de efectividad obtenida con diseños de antes y después puede ser discutible, y también se puede discutir si los desenlaces que escogieron los autores eran los más relevantes.

pues, si los beneficios valen lo que se invirtió en recursos; la intervención con el valor neto de la operación más positivo sería la que valdría la pena implementar. Sin embargo, la valuación explícita del desenlace en términos monetarios hace que esa alternativa de análisis no sea muy aceptada entre clínicos, ni, incluso, entre los pagadores de los servicios de salud; ni el valor monetario de una vida salvada ni el de un incremento en años de vida bien vividos son fácilmente conceptualizados ni medidos en FC[31].

En el análisis de costo-efectividad (CEA, por las siglas en inglés de ***cost-effectiveness analysis***), el consumo de recursos asociados a la utilización de un medicamento se mide como en los puntos anteriores, y se compara contra los desenlaces obtenidos, en términos fácilmente aceptados por clínicos, pacientes, y tomadores de decisiones (vidas salvadas, casos curados, casos prevenidos, hospitalizaciones evitadas, etc.). El resultado es, entonces, una razón de costos en términos monetarios sobre un desenlace en salud relevante, como pesos por vida salvada. Igualmente, dicha razón se compara contra un estándar del manejo, y se evalúa el costo incremental (en el margen) contra la efectividad incremental obtenida, o una razón de costo-efectividad incremental (ICER, por las siglas en inglés de ***incremental cost-effectiveness ratio***).

El CEA es, quizá, el tipo de estudio de evaluación económica para medicamentos más común presente en la literatura hoy día. Las limitaciones de este tipo de diseño incluyen la dificultad para la comparabilidad entre estudios (incluso, para la misma indicación), y, sobre todo, para su comparación contra otras enfermedades. Por ejemplo, un autor puede haber escogido hacer un CEA entre medicamentos para tratar la hipertensión, y haber escogido como desenlace costos por milímetro de mercurio (mmHg) disminuido; otro autor, costos por infarto prevenido, y otro autor, costos por evento de falla cardíaca prevenida. Otra dificultad involucra, como en el ejemplo anterior para el tratamiento de la hipertensión, cuando el efecto de los medicamentos para tratar la hipertensión impacta simultáneamente varios desenlaces, con afectación diferente de la calidad de vida y la supervivencia. Aunque en estos casos se puede hacer un ***análisis de costo-consecuencia***, con la estimación de los ICER para cada desenlace por separado, otra opción como los estudios de costo-utilidad es, quizás, más apropiada.

El ***análisis de costo-utilidad*** (CUA, por las siglas en inglés de ***cost-utility analysis***) sería la estrategia para elegir en la situación descrita con anterioridad, y la recomendada por varias guías, como la guía colombiana del Ministerio de Salud para elaborar análisis económicos asociados al desarrollo de guías de práctica clínica[32].

Algunos autores no distinguen entre CEA y CUA, dado que los principios de ambos son iguales: medir el consumo de recursos asociados a la utilización de un medicamento, como en los puntos anteriores, y compararlo contra los desenlaces obtenidos, en términos de un solo tipo de resultado, que, por lo general, son años de vida ajustados por calidad (QALY, por las siglas en inglés de *quality adjusted life years*).

Un QALY puede ser definido como un equivalente de un año de salud completa después de ajustar por lo deseable de ese año (debido a discapacidad, síntomas, u otros estados de enfermedad asociados). Como tal, un QALY se basa en la preferencia, o la "utilidad" que se le da (o que, de hecho, tiene) a un estado de salud.

En resumen, el CUA responde a una de las principales preocupaciones que se tienen respecto a los CEA, en el sentido de que las preferencias asociadas a diferentes desenlaces pueden ser cuantificadas y "traducidas" a una medida única; por lo general, un QALY. El CUA es particularmente útil cuando el medicamento bajo estudio tiene como uno de los principales desenlaces

cambios en la calidad de la vida, o cuando el investigador quiere comparar el resultado con los de otros estudios utilizando una "moneda común": el QALY. Cabe recalcar, sin embargo, que aún se debate si el QALY es o no la mejor medida; también, sobre la cuantificación de preferencias y utilidades, respecto a la decisión sobre descontar costos y desenlaces ocurridos en el futuro (por las preferencias sociales existentes de tener costos diferidos al futuro, pero desenlaces benéficos lo antes posible), y sobre la manera de cuantificar la incertidumbre que queda del análisis realizado[26].

La inmensa mayoría de los autores en farmacoepidemiología están de acuerdo en que una parte integral de los análisis farmacoeconómicos robustos incluye las estrategias de análisis de decisiones y un análisis de sensibilidad prolijo del impacto sobre los resultados que puedan producir cambios razonables en los estimativos más "blandos"[33]. Si bien se puede, en teoría, tener un estudio donde toda la información sobre los consumos de recursos y los desenlaces relevantes han sido identificados y cuantificados de forma válida y precisa, y, por ende, no necesitarse ningún ejercicio de modelamiento o de análisis de sensibilidad, todo ello rara vez es el caso. Muy a menudo, se necesitan las mencionadas estrategias, por la complejidad de las alternativas bajo estudio, de tener que proyectar costos y desenlaces fuera de las ventanas de tiempo de estudios clínicos, por dudas sobre la calidad de los datos de eficacia, o sobre qué tan generalizables son. Tal complejidad se debe, más o menos a menudo, a dudas sobre la identificación, la medición y la valuación de los recursos utilizados.

Este modelamiento puede ser muy complejo y lento para realizarse, y usualmente requiere recurso humano y programas especializados para su aplicación. Es relativamente usual que el modelamiento tome la forma de árboles de decisiones, en los cuales las intervenciones bajo estudio y sus consecuencias clínicas y económicas, así como la probabilidad de ocurrencia, se muestran explícitamente[34,35].

Estos modelos tienen, por lo general, cuatro componentes: los estados de enfermedad resultantes; las probabilidades de ocurrencia de ellos (o de transición entre ellos); los costos, los beneficios financieros y los efectos incurridos por los pacientes en cada estado; y los grupos de riesgo. Estos árboles, que suelen ser muy frondosos en sus ramas, no pueden mostrarse de forma completa en gráficas. Los datos para alimentar estos modelos varían sustancialmente, y, por ende, también lo hace la necesidad de evaluar cómo los cambios razonables en los supuestos asumidos en el estudio pueden afectar los resultados (proceso al que se llama ***análisis de sensibilidad***). Si a pesar de cambios razonables en los estimativos más "blandos" la dirección de los resultados no cambia de forma significativa, se dice que el resultado es "robusto".

Ejemplo de un análisis económico con medicamentos

Este análisis económico se desarrolló gracias a la financiación de Colciencias y del Ministerio de Salud de Colombia, dentro del marco de la Guía de Atención Integral para la Prevención, diagnóstico, atención y seguimiento de pacientes con hipertensión arterial, con el fin de ser utilizado como un insumo en las recomendaciones finales sobre tratamiento[36]. Por consenso con los financiadores, se decidió enfocar la evaluación económica en comparar el tratamiento inicial con monoterapia contra las principales intervenciones farmacológicas disponibles para el tratamiento de la hipertensión arterial (HTA) de leve a moderada, recién diagnosticada. Se pensó que este punto es muy relevante, pues las alternativas farmacológicas enfrentadas afectan desenlaces críticos (eventos cardiovasculares mayores), hay suficiente evidencia de elevada calidad respecto a la eficacia y la efectividad de las cinco principales alternativas farmacológicas como para informar el proceso de análisis económico, no hay clara evidencia *a priori* de la dominancia de una alternativa y hay elevada variabilidad en la práctica actual en la escogencia de las alternativas farmacológicas enfrentadas. Por ende, una evaluación económica podría ayudar a racionalizar sustancialmente el uso de recursos y contener costos.

Se consideró a pacientes (hombres y mujeres) de 45 años de edad o más, con diagnóstico reciente de HTA, que persistieran con HTA después de la dieta y de los cambios en su estilo de vida; fueron excluidos los pacientes con diagnóstico de enfermedad cardiovascular, falla cardíaca o diabetes.

Las intervenciones comparadas en el modelo son: no intervención (NI); diuréticos (D); bloqueadores de los canales del calcio/calcio-antagonistas (C); beta-bloqueadores (B); bloqueadores del receptor de angiotensina II

Continúa

Continuación

(BRA II) e inhibidores de la enzima convertidora de angiotensina (IECA) (A).

Desenlaces: La efectividad del tratamiento se midió por el impacto en la frecuencia de presentación de angina inestable (AI), infarto agudo de miocardio (IAM), falla cardíaca, ataque cerebro vascular (ACV) y muerte relacionada con enfermedad cardiovascular (ECV) como desenlaces primarios. Como desenlaces secundarios se incluyeron los efectos secundarios más comúnmente relacionados con el tratamiento antihipertensivo: por ejemplo, aparición de falla cardíaca o de diabetes tipo II. Como medida de utilidad al final de la vida, se tomaron los QALY. El principal desenlace económico fue el "costo incremental por QALY ganado" con cada tratamiento al final de la vida, comparado con la opción de menor costo (ICER).

Perspectiva: El análisis se consideró con la perspectiva del Sistema General de Seguridad Social en Salud (SGSSS) —pagador—. Por tanto, únicamente se consideran costos médicos directos y costos relacionados con la atención.

Horizonte de tiempo: Como la HTA tiene una morbilidad a largo plazo y consecuencias en la mortalidad, para el resultado primario del modelo se asumió un horizonte de tiempo hasta el final de la vida. La tasa de descuento usada para costos y desenlaces corresponde a una tasa constante del 3 % anual.

Información sobre efectividad y seguridad: Las probabilidades de ECV para la población de entre 45 y 55 años fueron estimadas a partir de los riesgos de eventos reportados en la literatura (ver "información en el modelo", más adelante), y del Departamento Administrativo Nacional de Estadística de Colombia (dane).

Estimación de Utilidades y QALY: Para Colombia, no se dispone de información válida sobre las preferencias de la población respecto a los diferentes estados de salud. La asignación de las utilidades asociadas a los diferentes estados de salud considerados se seleccionaron a través de la literatura.

Estimación y valoración del consumo de recursos: Para identificar y medir los recursos que se deberían costear en el caso de la atención hospitalaria de los desenlaces, se partió de la lista de recursos considerados en el modelo original desarrollado por el National Institute for Clinical Excelence del Reino Unido (NICE). Para identificar y medir los recursos que se deberían costear en el tratamiento ambulatorio, se buscó llegar a un consenso de expertos aplicando la metodología Delphi. Una vez identificados los recursos, para su valoración se usaron las tarifas del Instituto de Seguros Sociales (ISS) de 2001, ajustadas en el 30 % adicional al valor definido.

Estimación del costo de los medicamentos: Para estimar los costos de los medicamentos, se tomaron las bases de datos del Sistema Integral de Información de la Protección Social (SISPRO), y, específicamente, del Sistema de información de Precios de Medicamentos (SISMED), del Ministerio de la Protección Social.

Estructura y consideraciones del modelo: Se consideró un modelo markoviano para evaluar los costos incrementales y los desenlaces incrementales del tratamiento a largo plazo de la hta de leve a moderada. En el modelo, todos los pacientes hipertensos empiezan en un estado libre de desenlaces. Los ciclos de tránsito entre los distintos desenlaces fueron modelados cada seis meses, a lo largo de los cuales una proporción importante de los pacientes hipertensos se mantendrían libres de eventos. Todos los pacientes terminan en el modelo en un estado de muerte. La **figura 24.1** muestra una representación esquemática del modelo.

Variabilidad e incertidumbre en la información: Para evaluar el impacto de la incertidumbre en los supuestos usados en el análisis del caso de referencia, se hicieron análisis de sensibilidad univariados de tipo determinístico. Para el análisis de sensibilidad probabilístico, se tomó como base el caso de referencia. El plano de costo-efectividad se hizo a través de un proceso de simulación de 1000 repeticiones, en el que se variaron los valores del modelo. Para estimar los valores de los parámetros de probabilidad, se decidió generar los valores de los parámetros a través de una distribución Beta. Para las variables de costos, los valores de los parámetros se generaron a través de la distribución Gamma. Para las variables de utilidad, también se decidió generar los valores a través de una distribución Beta.

Resultados caso base: Los costos y los desenlaces medidos en QALY de las estrategias bajo comparación se muestran en la **tabla 24.4**, para el caso base (hombre y mujer de 65 años, el 2 % de riesgo anual de enfermedad cerebrovascular (CV), el 1 % de falla cardíaca, el 1,1 % de diabetes, con el costo promedio ponderado del grupo de medicamentos, tasa de descuento del 3 %, y seguimiento por 20 años. Este análisis muestra que los diuréticos son menos costosos y más efectivos que la no intervención y que los betabloqueadores. Los calcioantagonistas tienen un incremento mínimo en QALY respecto a los diuréticos, pero su costo incremental es muy alto. Llama la atención que tanto el grupo de BRA II / IECA como el de diurético tienen una efectividad (QALY) indistinguible, pero el costo asociado a ARA II / IECA (para ARA II = antagonistas de los receptores de angiotensina II) es mucho mayor. La **figura 24.2** trae el plano de costo-efectividad de los grupos de medicamentos contra la estrategia de no intervención.

Análisis de sensibilidad probabilístico: Dado el escenario basal en el que el diurético fue la opción de menor costo, se decide mostrar en este análisis la distribución de la costo-efectividad incremental contra la opción de no intervención. Como se muestra en la **figura 24.3**, el tratamiento antihipertensivo (en este caso, con diurético) será dominante respecto a la no intervención en la mayoría de los escenarios, y que en los escenarios donde no es dominante, sí es claramente costo-efectivo.

Tabla 24.4 ICER respecto al diurético, para el caso de referencia (*), para hombres y mujeres

	Costo (COP) por persona	Efecto (QALY por persona)	ICER* (costo por QALY)
Hombres			
D	6 498 624	9,24	Opción de menor costo
A	13 178 919	9,24	6 680 295; QALY igual
C	27 774 098	9,26	1 063 773 700
B	8 563 374	9,03	Dominado
NI	**6 769 607**	**8,80**	**Dominado**
Mujeres			
D	6 503 235	9,49	Opción de menor costo
A	13 324 399	9,49	6 821 164; QALY igual
C	28 340 040	9,51	1 091 840 250
B	8 626 771	9,27	Dominado
NI	**6 776 312**	**9,03**	**Dominado**

* Hombre y mujer de 65 años, 2 % de riesgo anual de enfermedad CV, 1 % de falla cardíaca, 1,1 % de diabetes, con el costo promedio ponderado del grupo de medicamentos, tasa de descuento del 3 %, seguimiento por 20 años, y asumiendo que el 20 % de los sujetos serán intolerantes a IECA y necesitarán un ARA.

D: diurético; A: BRA II / IECA; C: calcioantagonistas; B: beta-bloqueadores; NI: no intervención; QALY: año de vida ajustado por calidad (*quality adjusted life year*).

Fuente: Rojas, et al.[36].

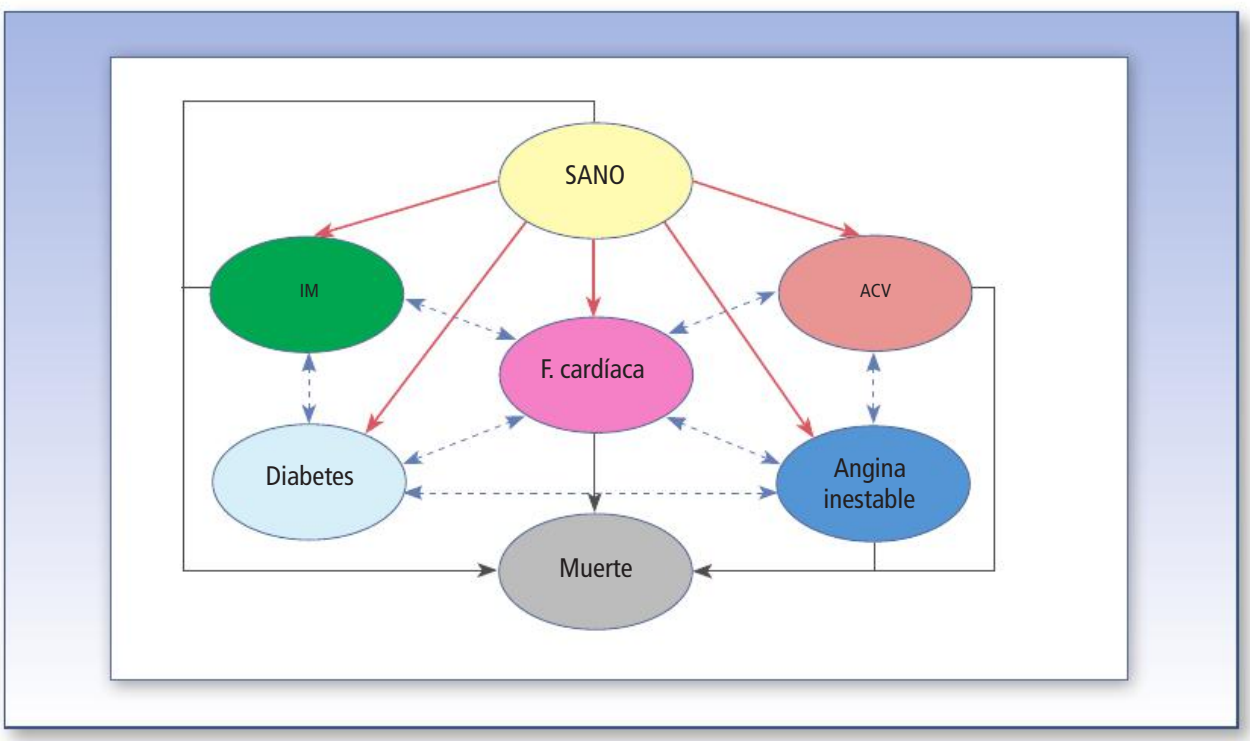


Figura 24.1 Estructura del modelo de Markov para evaluar la razón de costo-efectividad de los medicamentos para el tratamiento de la HTA. IM: infarto de miocardio; ACV: ataque cerebrovascular; FCardíaca: falla cardíaca.

Fuente: Esquema proporcionado por el grupo desarrollador del modelo en el NICE.

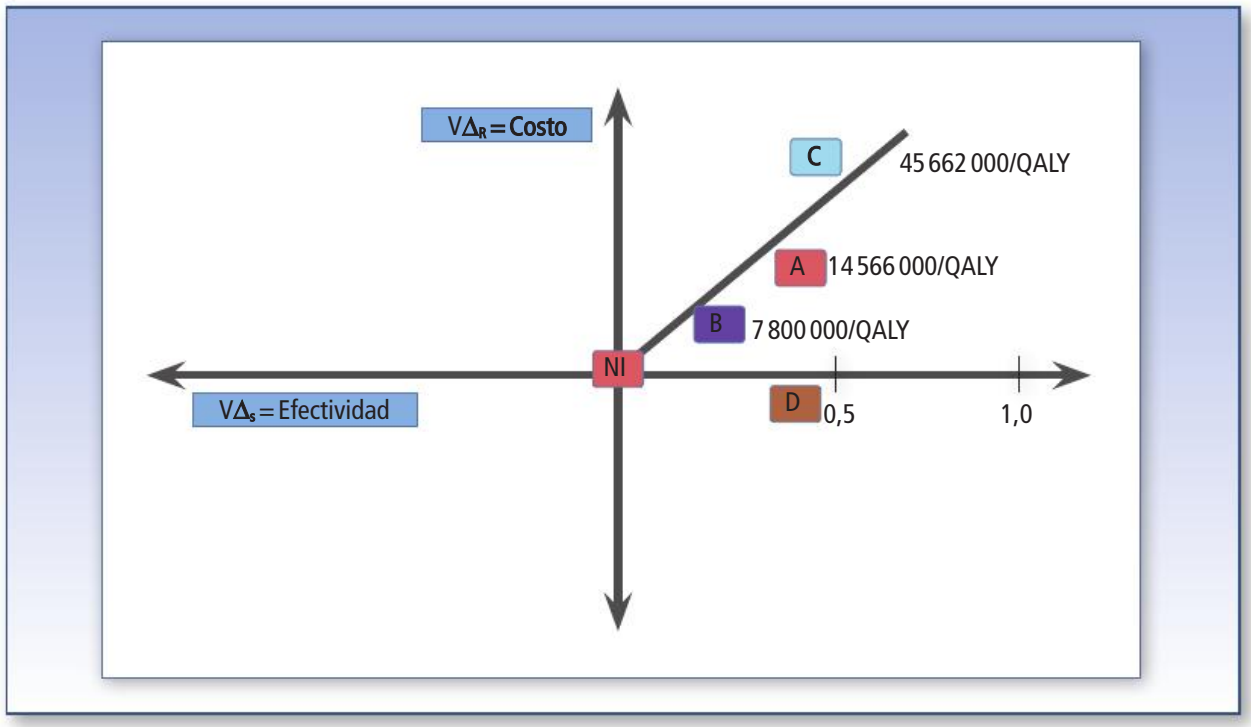


Figura 24-2 **Resultados del análisis de costo-efectividad respecto a NI, para el caso de referencia (*) para hombres y mujeres.**

* Hombre y mujer de 65 años, 2 % de riesgo anual de enfermedad CV, 1 % de falla cardíaca, 1,1 % de diabetes, con el costo promedio ponderado del grupo de medicamentos, tasa de descuento del 3 %, seguimiento por 20 años, y asumiendo que el 20 % de los sujetos serán intolerantes a IECA y necesitarán un ARA. **D: diurético; A: A:BRA II / IECA; C: calcioantagonistas; B: beta-bloqueadores; NI: no intervención; QALY:** año de vida ajustado por calidad (*qualitiy adjusted life year*).

Fuente: Rojas, et al.[36].

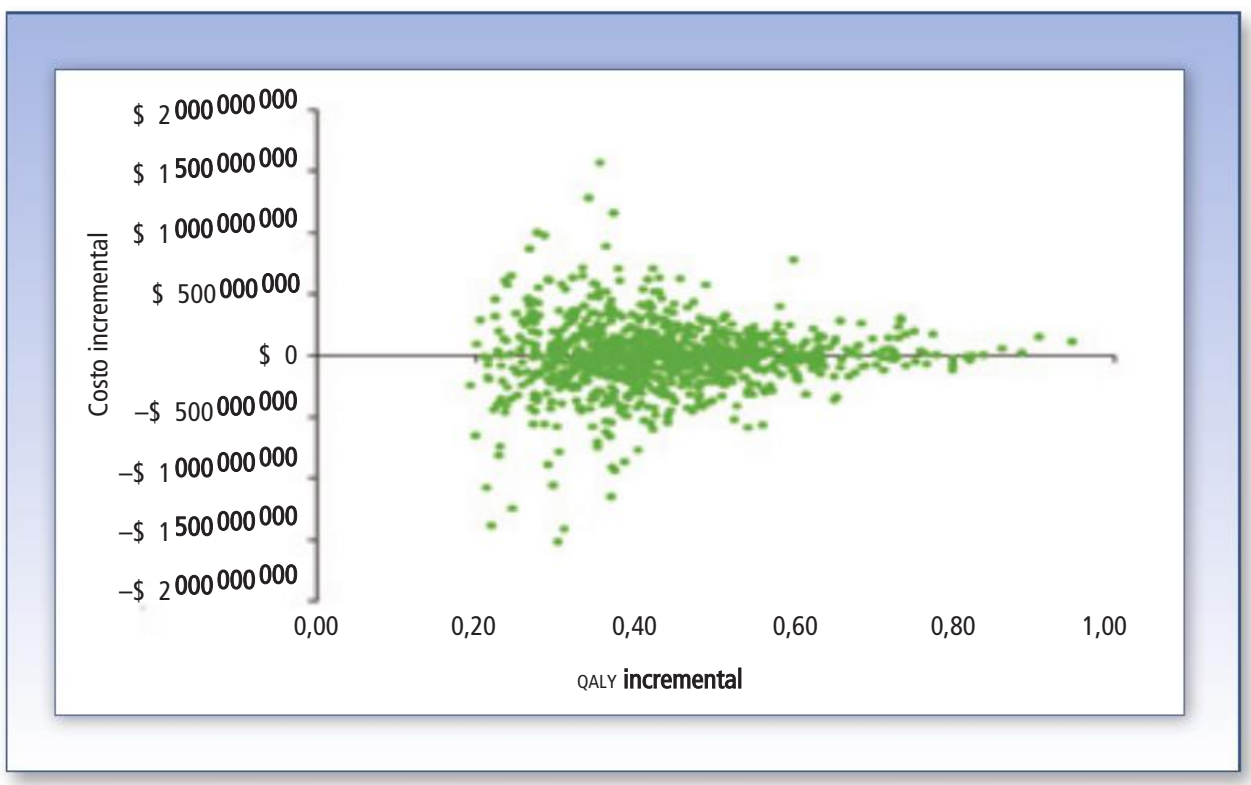


Figura 24.3 **Análisis de sensibilidad probabilístico de la relación de costo-efectividad incremental de la terapia con diurético** contra la opción de **NI . QALY**: año de vida ajustado por calidad (*quality adjusted life year*).

Fuente: Rojas, et al.[36].

Referencias

1. Strom BL, Tugwell P. Pharmacoepidemiology: Current status, prospects, and problems. Ann Intern Med. 1990;113:179-81.
2. Kabiri M, Haswinski AB, Roberts MS, et al. The changing burden of hepatitis C virus infection in the United States: Model-Based predictions. Ann Intern Med. 2014;161:170-80.
3. Briggs A, Levy AR. Pharmacoeconomics and pharmacoepidemiology: curious bedfellows or a match made in heaven? Pharmacoeconomics. 2006;24: 1079-86.
4. Ciociola AA, Cohen LB, Kulkarni P, et al. How drugs are developed and approved by the FDA: Current process and future directions. Am J Gastroenterol. 2014;109:620-3.
5. Pronker ES , Weenen TC, Commandeur HR, et al. The gold industry standard for risk and cost of drug and vaccine development revisited. Vaccine. 2011;29:5846-9.
6. Strom BL. What is pharmacoepidemiology? En: Strom B, Kimmel S, Hennessy S, editores. Textbook of pharmacoepidemiology. 2nd ed. Oxford: Editorial Wiley Blackwell; 2013.
7. Lis Y, Roberts MH, Kambles, et al. Comparisons of Food and Drug Administration and European Medicines Agency Risk Management Implementation for Recent Pharmaceutical Approvals. Value Health. 2012;15:1108-18.
8. Wettermark B. The intriguing future of pharmacoepidemiology. Eur J Clin Pharmacol. 2013;69:43-51.
9. Scurti V, Romero M, Tognoni G. A plea for a more epidemiological and patient-oriented pharmacovigilance. Eur J Clin Pharmacol. 2012;68:11-9.
10. Schneeweiss S, Suissa S. Advanced approaches to controlling confounding in pharmacoepidemiologic studes. En: Strom B, Kimmel S, Hennessy S, editores. Textbook of Pharmacoepidemiology. 2nd ed. Oxford: Editorial Wiley Blackwell; 2013
11. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol. 2008;167:492-9.
12. Hanley J, Dendukuri N. Efficient sampling approaches to address confounding in database studies. Stat Methods Med Res. 2009;18:81-105.
13. Avorn J. The promise of pharmacoepidemiology in helping clinicians assess drug risk. Circulation. 2013;128:745-8.
14. Strom B, Kimmel SE, Hennessy S. Textbook of pharmacoepidemiology. 2nd ed. Oxford: John Wiley & Sons Ltd; 2013.
15. Hennessy S. Use of health care databases in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98:311-3.
16. Hall GC, Sauer B, Bourke A, et al. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. 2012;21:1-10.
17. Takahashi Y, Nishida Y, Asai S. Utilization of health care databases for pharmacoepidemiology. Eur J Clin Pharmacol. 2012;68:123-9.
18. Etminan M. Pharmacoepidemiology II: The nested case-control study—a novel approach in pharmacoepidemiologic research. Pharmacother J Hum Pharmacol Drug Ther. 2004;24:1105-9.
19. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98:253-9.
20. Bigby M, Jick S, Jick H, et al. Drug-induced cutaneous reactions. JAMA. 1986;256:3358-63.
21. Nordmann S, Biard L, Ravaud P, et al. Case-only designs in pharmacoepidemiology: a systematic review. PloS One. 2012;7:e49444.
22. Consiglio GP, Burden AM, Maclure M, et al. Case-crossover study design in pharmacoepidemiology: systematic review and recommendations. Pharmacoepidemiol Drug Saf. 2013;22:1146-53.
23. Hallas J, Pottegard A. Use of self-controlled designs in pharmaco-epidemiology. J Intern Med. 2014;275:581-9.
24. Dennis R, Pinto D. Farmacoeconomía. Acta Medica Colomb. 2006;31:53-5.
25. Pinto D. Estudios de análisis económico. En: Epidemiologia clínica. Ruiz A, Morillo L, Eds. Bogotá: Editorial Médica Panamericana; 2004.
26. Schulman KA, Glick HA, Reed SD. Pharmacoeconomics: Economic evaluation of Pharmaceuticals. En: Strom B, Kimmel SE, Hennessy S. Textbook of Pharmacoepidemiology. 2nd ed. Oxford: John Wiley & Sons Ltd; 2013. Capítulo 17.
27. Yong JH, Beca J, Hoch JS. The evaluation and use of economic evidence to inform cancer drug reimbursement decisions in Canada. Pharmacoeconomics. 2013;31:229-36.
28. Dennis R. Editorial: Evaluación económica de la denervación simpática renal: simplicidad, todavía no. Rev Colomb Cardiol. 2014;21:152-3.
29. Drummond M, Sculpher M, Torrance G, et al. Methods for the economic evaluation of health care programmes. 3rd ed. Nueva York: Oxford University Press; 2005.
30. Woodward RS, Medoff G, Smith MD, et al. Anti-

biotic cost savings from formulary restrictions and physician monitoring in a medical-school-affiliated hospital. Am J Med. 1987;817-23.

31. Bodrogi J, Kaló Z. Principles of pharmacoeconomics and their impact on strategic imperatives of pharmaceutical research and development. Br J Pharmacol. 2010;159:1367-73.
32. Colombia,Ministerio de la Protección Social, COLCIENCIAS, et al. Guía metodológica para la elaboración de Guías de Atención Integral en el sistema General de Seguridad Social en Salud colombiano. Bogotá, Minsalud; 2010.
33. Eddy D, Holligworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices Task force-7. Med Decis Making. 2012;32:733-43.
34. Ademi Z, Kim H, Zomer E, et al. Overview of pharmacoeconomic modelling methods. Br J Clin Pharmacol. 2013;75:944-50.
35. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force 2. Value Heal J Int Soc Pharmacoeconomics Outcomes Res. 2012;15:804-11.
36. Rojas MX, Dennis R, Rondón M, et al. Evaluación económica de las principales intervenciones farmacológicas disponibles como monoterapia para el tratamiento de la hipertensión arterial leve a moderada recién diagnosticada. En: Prevención, Diagnóstico, Atención integral y Seguimiento de pacientes con Hipertensión Arterial. Convocatoria Colciencias y Ministerio de la Protección Social, 2011-2012.

CAPÍTULO

25 Evaluaciones de tecnología

INTRODUCCIÓN

Durante las últimas tres décadas, se ha vivido un crecimiento exponencial de la industria de la salud que ha llenado el mercado de nuevas tecnologías, ya sea como resultado del avance de disciplinas clásicas (bioquímica y genética) o por el desarrollo de nuevas disciplinas (ciencias de la información)[1]. Como consecuencia, los sistemas de salud de los países desarrollados enfrentan hace ya varios años la problemática de decidir cuáles de las tecnologías disponibles usar o incluir en los planes de cobertura del Estado, y en qué casos o poblaciones dichas tecnologías se deben usar prioritariamente, problemática que comparten los sistemas de salud de todo el mundo[2]. La necesidad de tener una base bien documentada acerca de la efectividad y la seguridad de las tecnologías existentes y de las emergentes, para la planificación y la toma de decisiones adecuadas en salud ha crecido casi con la misma rapidez con la que surgen nuevas tecnologías.

Las decisiones sobre el uso de las tecnologías se hacen en todos los niveles del sistema de atención de salud, y a menudo incluyen no solo unificar la información médica compleja, sino también, información relacionada con el paciente, con la organización y con el impacto económico que tendría el uso de la tecnología, en un contexto donde puede haber cuestionamientos éticos. Tales decisiones pueden estar relacionadas con la adquisición, la contratación pública, la financiación o el uso apropiado de las tecnologías de la salud, así como con la desinversión en tecnologías obsoletas o ineficaces.

Dadas las implicaciones en el capital de salud, así como en el capital social de un país, las decisiones sobre el uso de tecnologías en salud deben tener una fundamentación basada en evidencia, donde todas las circunstancias y las consecuencias relevantes estén sistemáticamente ilustradas por métodos científicos[2]. Por lo anterior, recopilar la información necesaria y presentarla a los tomadores de decisiones es una tarea que depende en grado sumo de la interacción, la división del trabajo y la cooperación entre los profesionales del sistema de salud, los conocedores de la investigación y los responsables políticos o sus funcionarios[3].

La ***evaluación de tecnologías en salud*** (**ET**) surgió como respuesta a la necesidad de información para fundamentar decisiones, y se constituyó en una forma de transferencia del conocimiento, pues se ocupa de organizar la información de varios aspectos relacionados con los efectos, las implicaciones y las consecuencias derivadas del uso de las tecnologías y

presentarla a los encargados de tomar decisiones en salud[4].

Si bien la ET se halla conectada con una aproximación sistemática, basada en la investigación, no hay un método único o un campo de investigación delimitado a ese respecto. Más bien, la ET se caracteriza por ser interdisciplinaria y con un enfoque versátil, donde la base científica se construye a partir de teorías y métodos de un número de diferentes disciplinas de investigación[5]. Su función es proveer información válida y necesaria para soportar cambios en las políticas y las prácticas de cuidado de la salud, más que la producción de conocimiento *per se*[6].

Las primeras instituciones u órganos dedicados a evaluar tecnologías de salud se establecieron en Francia y España a principios de la década de los ochenta, y la primera Agencia Nacional de Evaluación de Tecnologías en Salud hizo lo propio en Suecia, en 1987[7]. Desde entonces, el crecimiento en el número de instituciones involucradas en la evaluación de tecnologías ha llevado a que hoy se cuente con experimentadas agencias en varios países del mundo (en su mayoría, de Europa y de Norteamérica)[7]. Dichas agencias, así como las redes o las asociaciones que las agrupan, han establecido metodologías y marcos de trabajo que guían los procesos para llevar a cabo ET dentro de estándares de calidad. Entre las asociaciones que han hecho mayores aportes al desarrollo de ET se pueden citar: International Network of Agencies for Health Technology Assessment (INAHTA); European Network for Health Technology Assessment (EUNETHTA); Health Technology Assesment International (HTAI) y The International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Este capítulo busca presentar al lector una síntesis de los métodos y los procesos que guían la conducción de ET de alta calidad. Se basa en una revisión rigurosa de los manuales[4-11], las directrices y los modelos[4] propuestos por las más destacadas agencias de ET y las organizaciones ya mencionadas, con el fin de identificar procesos y lineamientos comunes para el proceso de ET.

El capítulo está dirigido a las personas que participan en la planificación y la ejecución de proyectos de ET o que buscan hacer ET, tales como profesionales de la salud, responsables de decisiones políticas y administrativas, grupos de interés, investigadores y otras personas que quieran adoptar una actitud crítica hacia las ET, incluyendo a profesores y estudiantes.

La interconexión compleja de la ciencia, la sociedad, la formulación de políticas y los problemas relacionados con la toma de decisiones en salud constituyen los temas centrales de la ET.

CONCEPTOS BÁSICOS

Es pertinente aclarar aquí algunos conceptos referentes a las tecnologías en salud, que facilitarán la comprensión de los contenidos del presente capítulo:

Tecnología

Se refiere a la aplicación práctica del conocimiento. Tecnología "[…] es el conjunto de conocimientos técnicos, científicamente ordenados, que permiten diseñar y crear bienes y servicios que facilitan la adaptación al medio ambiente y satisfacer tanto las necesidades esenciales como los deseos de la humanidad"[12]. A pesar de dicha definición, la palabra "tecnología" es asociada, generalmente, a desarrollos en electrónica o en dispositivos mecánicos; pero en salud, el término tiene una connotación más amplia, pues la aplicación práctica del conocimiento en salud encuentra muchos espacios, que van desde la biología hasta lo organizacional y lo social.

Tecnologías en salud

Se refiere a cualquier intervención que pueda utilizarse para promover la salud, con el fin de prevenir, diagnosticar o tratar una enfermedad, o para la rehabilitación o la atención a largo plazo. Esto incluye los productos farmacéuticos, los productos sanitarios, los procedimientos y los sistemas organizativos utilizados en la atención de la salud[13].

Goodman propone describir las tecnologías para el cuidado de la salud de tres formas: según 1) su naturaleza, 2) su propósito y 3) su estado de difusión[14]:

- Según su naturaleza: En salud, se identifican tecnologías de diferente naturaleza que pueden agruparse en cuatro categorías principales: 1) las drogas o medicamentos (por ejemplo: analgésicos, antibióticos, antiinfla-

matorios, etc.); 2) los biológicos (por ejemplo: vacunas, derivados sanguíneos, células y recursos genéticos); 3) los dispositivos, los equipos y los suministros (por ejemplo: marcadores cardíacos, prótesis, equipos de diagnóstico, endoscopios, etc.); 4) los procedimientos médicos y quirúrgicos (por ejemplo: la psicoterapia, las técnicas quirúrgicas, la aplicación de ***stents***, los programas de salud —clínicas de seguimiento, formas de entrega del cuidado, etc.—)[14]. Hay también algunas ***tecnologías combinadas***, que incluyen drogas, dispositivos o tecnologías de otras categorías; algunos ejemplos son las tecnologías para entrega de medicamentos (conocidas comúnmente como ***bombas***), que constan, a su vez, de dispositivos implantables que liberan drogas o medicamentos.

- Según su propósito o su aplicación: El propósito que tienen las tecnologías en salud puede variar según el momento del curso clínico de la enfermedad en el cual son aplicadas, así que las tecnologías en salud pueden ser para: a) prevención (reducir el riesgo de ocurrencia de una enfermedad o limitar su evolución; por ejemplo: programas de inmunizaciones, programas de control de infecciones, intervenciones profilácticas como la terapia con flúor); b) tamización (identificación temprana de una enfermedad, una anormalidad o una condición de riesgo en la población; por ejemplo: la medición de la presión arterial o del colesterol, o las citologías o las mamografías); c) diagnóstico (identificación o diferenciación de una enfermedad en personas con síntomas sugestivos de padecerla; por ejemplo: electrocardiogramas para confirmar un infarto de miocardio, pruebas serológicas para el virus de la inmunodeficiencia humana [VIH], cultivos de laboratorio para infecciones); d) tratamiento (mejorar o mantener la salud, evitar el deterioro; por ejemplo: terapia antiviral, antibiótica, puentes coronarios, dispositivos para el control y el manejo del dolor, etc.); e) rehabilitación (restaurar, mantener o mejorar la funcionalidad y el bienestar de las personas física o psicológicamente discapacitadas; por ejemplo: programas de ejercicio postinfarto, dispositivos para mejorar la audición en pacientes con daños auditivos, programas de control vesical, etc.)[14]. Es claro que muchas de las tecnologías pueden tener más de una aplicación y ser usadas en la tamización y el diagnóstico, o en el diagnóstico y en el tratamiento (por ejemplo, la angiografía coronaria), así como en la prevención.
- Según su estado de difusión: Se puede decir que las tecnologías tienen un "ciclo de vida", y que pueden ser evaluadas en diferentes momentos de ese ciclo: en el comienzo de su desarrollo, en el inicio de su difusión, en el curso de su aplicación o en el momento en el que van perdiendo vigencia (son reemplazables por nuevas tecnologías que aportan mayores beneficios en salud). De acuerdo con ello, Goodman[14] propone que, según su estado de difusión, se clasifiquen las tecnologías en salud en: ***experimentales*** (se están probando en animales o en modelos); ***de investigación clínica*** (se inicia su evaluación para uso en seres humanos con una condición o una indicación particular); ***establecidas*** (aprobadas e incluidas en la práctica general para un uso particular o una indicación); ***obsoletas*** (han mostrado no ser tan efectivas, o pueden ser dañinas y tienen el potencial de ser reemplazadas por otras tecnologías de su clase o nuevas, que presenten mayores beneficios en salud)[14].

Actores o partes interesadas

Se refiere a quienes pueden afectar o, de hecho, son afectados, por las actividades de una empresa[15]. Se puede decir son los públicos interesados que deben considerarse un elemento esencial en la planificación estratégica de cualquier actividad de negocios.

Dado el impacto de las ET en la provisión de cuidado de la salud, son muchos los interesados en el proceso que siguen los gobiernos para definir cuáles tecnologías amerita evaluar[6]. Entre ellos se cuentan la industria farmacéutica, los financiadores de investigación, los proveedores de salud y de servicios sociales y los pacientes que pueden recibir el beneficio de nuevas tecnologías o de intervenciones que las incluyan[6].

DEFINICIÓN

La evaluación de tecnologías (ET) en salud ha sido definida como "una forma de aplicación de la investigación para políticas, que examina sistemáticamente las consecuencias a corto y largo plazo de la aplicación de una tecnología, en

términos de salud y uso de recursos"[16]. La ET se ocupa de las consecuencias médicas, organizacionales, económicas y sociales de implementar tecnologías de salud o intervenciones dentro del sistema de salud[17].

El Centro Danés para Evaluación de Tecnologías define la ET como "la evaluación basada en la investigación y orientada a la utilización del conocimiento relevante disponible acerca de los problemas asociados con el uso de una tecnología en relación con la salud y la enfermedad. Por virtud de estos métodos, basados en evidencia pero también en términos de su intención de apoyar la toma de decisiones, la ET está relacionada con la planeación administración y gestión de los servicios de salud"[5].

ORIENTACIÓN Y TIPOS DE INFORMES DE ET

Existen tres orientaciones que se le pueden dar a la ET[14]:

- Evaluación orientada a la tecnología: Intenta determinar las características y el impacto de una tecnología en un problema en particular (por ejemplo, si se quiere determinar el impacto clínico, social, organizacional y económico del uso de implantes cocleares en la población de niños con sordera).
- Evaluación orientada al problema o condición: Busca soluciones o estrategias para el manejo y el control de un determinado problema de salud para el cual existen tecnologías alternativas o complementarias.
- Evaluación orientada a un proyecto: Se caracteriza por tener una aplicación local, como el uso de una tecnología en una institución o un programa específico (por ejemplo, el Hospital de la Santa Creu i Sant Pau, en Barcelona, como parte del proyecto MAPAC, hace periódicamente ET sobre las diferentes tecnologías que se utilizan en los servicios clínicos, con el fin de mejorar la adecuación de la practica asistencial en el hospital y regular los recursos).

DOMINIOS O ASPECTOS QUE INCLUYE LA ET

De acuerdo con las definiciones presentadas anteriormente, la ET compromete análisis y evaluaciones de varias áreas en las cuales el uso de la tecnología puede traer consecuencias. Estas áreas se pueden agrupar en cuatro elementos: 1) *la tecnología*, 2) *el paciente*, 3) *la organización* y 4) *la economía*. Estos cuatro elementos se superponen en cierta medida: por ejemplo, la ética, que es tradicionalmente atribuida al paciente o a la sociedad, no se puede separar del efecto de la tecnología ni de las consecuencias de su uso en la organización o en la economía; por tanto, es un elemento que hace parte del marco de análisis de los otros elementos[18]. Varios autores y organizaciones coinciden en incluir los mencionados elementos dentro de los siguientes *dominios*, los cuales deben ser considerados en la evaluación de una tecnología y delimitan su estructura:

- Problema de salud y uso actual de la tecnología.
- Descripción de la tecnología y sus características.
- Efectividad clínica y seguridad de la tecnología.
- Impacto económico (costos y costo-efectividad).
- Aspectos éticos.
- Aspectos sociales.
- Aspectos organizacionales.
- Aspectos legales.

En el *HTA CoreModel Handbook*, la EUNETHTA[19] propone una serie de interrogantes para considerar por parte de quienes hacen las ET, sobre cada uno de los mencionados dominios, y que sirven como guía para definir los contenidos más relevantes de la evaluación, de acuerdo con cada tecnología en particular[18]. Algunos de dichos interrogantes se presentan en la **tabla 25.1**.

Problema de salud y uso actual de la tecnología

Este dominio busca mostrar la importancia del problema de salud al cual se aplica la tecnología bajo evaluación, en términos de la población objetivo de la tecnología, la epidemiología, la carga de la enfermedad y las implicaciones a escala individual y para la sociedad. Esta información es particularmente útil para los encargados de tomar decisiones en el momento de definir prioridades en salud respecto a oferta de servicios, inclusión de medicamentos en planes de beneficios o qué tecnologías evaluar primero[5,19].

A mayor cantidad de personas afectadas, mayor la probabilidad de que una tecnología que aborde el problema sea de evaluación prioritaria.

Descripción de la tecnología y sus comparadores

En este dominio se busca presentar información relevante sobre la tecnología en evaluación y otras tecnologías existentes para la misma condición, buscando identificar los aspectos (pros y contras) de la nueva tecnología a los cuales se debe prestar mayor atención en la evaluación. Es pertinente definir el estado del ciclo de vida en el que se encuentran la tecnología y sus comparadores, para qué se utilizan, cómo actúan, quiénes los utilizan, qué nivel de complejidad se requiere para su aplicación, los requerimientos y los patrones para su uso (infraestructura, tipo y nivel de entrenamiento del personal, etc.), su disponibilidad en el contexto donde se da la ET y el estatus regulatorio en los ámbitos local e internacional. La **tabla 25.2** presenta algunos interrogantes que deben considerar los desarrolladores de la ET sobre la tecnología y su comparador[5,19].

Tabla 25.1a Preguntas que la ET debe responder sobre cada aspecto que involucra la toma de decisiones en salud

La evaluación de una tecnología deberá dar respuesta a los siguientes interrogantes	
Respecto al problema de salud y al uso actual de la tecnología:	**Acerca de la efectividad de la tecnología, según como sea pertinente para cada tecnología en evaluación:**
1. ¿Para qué condiciones de salud, y con qué propósito se utiliza la tecnología?	1. ¿Cuál es el efecto de la tecnología en las funciones mentales, sensoriales y de dolor, de voz y lenguaje, cardíacas, respiratorias e inmunes, genitourinarias y reproductivas, relacionadas con el movimiento y la piel del paciente?
2. ¿Cuál es la población objetivo (en términos de su enfermedad o su condición) para la evaluación actual de la tecnología?	2. ¿Cuál es el efecto de la tecnología en la calidad de vida relacionada con la salud?
3. ¿Cuáles son los factores de riesgo conocidos para la enfermedad o la condición de salud?	3. ¿Cuál es el efecto de la tecnología en calidad de vida según enfermedades específicas?
4. ¿Cuál es el curso natural de la enfermedad o la condición de salud?	4. ¿Cuál es el efecto de la tecnología en la capacidad de trabajo?
5. ¿Cuáles son los síntomas y la carga de enfermedad para el paciente en las diferentes etapas de su enfermedad?	5. ¿Cuál es el efecto de la tecnología, en contraste con las condiciones de vida previa?
6. ¿Cuáles son las consecuencias de la enfermedad o la condición de salud para la sociedad (es decir, la carga de la enfermedad)?	6. ¿Cómo se ven afectadas las actividades de la vida diaria con el uso de la tecnología?
7. ¿Qué aspectos de las consecuencias-carga de enfermedad son el objetivo de la tecnología?	7. ¿Cuáles son los beneficios y los daños generales de la tecnología en desenlaces de salud?
8. ¿Cuántas personas pertenecen a la población objetivo?	8. ¿Cuál es el efecto benéfico esperado de la intervención sobre la mortalidad general?
9. ¿Cuáles son las diferencias en el manejo para las diferentes etapas de la enfermedad o la condición de salud?	9. ¿Cuál es el efecto benéfico esperado de la intervención sobre la mortalidad específica relacionada con la condición?
10. ¿Cómo se diagnostica actualmente la condición de salud, según las guías de recomendaciones publicadas y a la práctica (si la et es de una prueba o examen diagnóstico)?	10. ¿Cómo puede la tecnología afectar los síntomas y los hallazgos (severidad, frecuencia) de la condición objetivo?
11. ¿Cómo se maneja actualmente la enfermedad o la condición de salud según las guías de recomendaciones publicadas y según la práctica?	11. ¿Cómo puede la tecnología afectar la progresión o la recurrencia de la condición objetivo?
12. ¿Cuál población es elegible para la tecnología, y con qué bases?	12. ¿Cómo puede la tecnología modificar la necesidad de hospitalización?
13. ¿Cuáles son las otras alternativas típicas o comunes de la tecnología actual?	
14. ¿Qué tanto se utiliza la tecnología actualmente, y qué tanto se utilizará en el futuro?	
15. ¿Qué tipo de variaciones en el uso de la tecnología se encuentran entre países, regiones y escenarios?	

Fuente: HTA Core *Model Handbook*[19].

Tabla 25.1b Preguntas que la ET debe responder sobre cada aspecto que involucra la toma de decisiones en salud

Respecto a la tecnología y su comparador:	Acerca de la seguridad de la tecnología:
1. ¿Quién fabrica la tecnología?	1. ¿Qué tipo de daños puede causar al paciente el **uso de la tecnología? ¿Cuáles son la incidencia, la** severidad y la duración de estos daños?
2. ¿Qué características tiene esta tecnología?	2. ¿Están los daños relacionados con la dosificación o la **frecuencia de aplicación de la tecnología?**
3. ¿En qué fase de desarrollo e implementación está la **tecnología?**	3. ¿Cómo cambian a lo largo del tiempo y en distintos **escenarios la severidad o la frecuencia de los daños?**
4. ¿Cuáles son la indicación aprobada y el beneficio que se **alega de la tecnología?**	4. ¿Hay grupos susceptibles de pacientes que sean más **propensos a sufrir daños por el uso de la tecnología?**
5. ¿Cuáles son las ventajas y las desventajas conocidas en el **tratamiento de la indicación de interés?**	5. ¿Hay cuestiones especiales en el uso de la tecnología **que puedan aumentar el riesgo de eventos dañinos?**
6. ¿Cuáles son las tecnologías que se usan actualmente para la **indicación de interés?**	6. ¿Qué tan segura es la tecnología en relación con la **que se compara?**
7. ¿Cómo se espera que haya una mejoría sobre las **tecnologías previas o existentes usadas para el mismo** problema de salud?	7. ¿Qué tipo de daños ocupacionales pueden ocurrir **cuando se usa la tecnología?**
8. ¿En qué contexto y en qué nivel de atención son utilizados **la tecnología y su comparador?**	8. ¿Qué tipo de riesgos existen, y qué daños pueden **ocurrir al público y al medio ambiente, cuando se usa** la tecnología?
9. ¿Qué clase de procesos de calificación y de garantía de **calidad son necesarios para el uso o el mantenimiento de la** tecnología?	9. ¿Cómo varía el perfil de seguridad de la tecnología **entre las diferentes generaciones, las versiones** aprobadas y los productos?
10. ¿Qué tipo de capacitación y de información son necesarios **para el personal/cuidador que utiliza la tecnología?**	10. ¿Cómo se pueden reducir los riesgos para los **pacientes (incluyendo aspectos dependientes de la** tecnología, del usuario y del paciente)?
11. ¿Qué equipos y qué suministros se necesitan para usar la **tecnología y el comparador?**	
12. ¿Qué tipo de registros son necesarios para supervisar el uso **de la tecnología y el comparador?**	

Fuente: HTA Core *Model Handbook*[19].

Tabla 25.1c Preguntas que la ET debe responder sobre cada aspecto que involucra la toma de decisiones en salud

Respecto a las consideraciones éticas relacionadas con el uso de la tecnología:	Respecto al consumo de recursos y costo-efectividad asociados a la tecnología en evaluación:
1. ¿Quién es el candidato correcto para la tecnología?	1. ¿Qué tipo de recursos son utilizados cuando se **entrega la tecnología evaluada y sus comparadores** (identificación de la utilización de recursos)?
2. ¿Está la tecnología destinada a mejorar la salud, la calidad **de vida relacionada con la salud, la calidad de vida o** la supervivencia, en comparación con las tecnologías alternativas?	2. ¿Cuáles son los resultados relacionados con la salud **(medidos o estimados) de la tecnología evaluada y de** sus comparadores?
3. ¿Puede la tecnología, de alguna manera, perjudicar a **pacientes individuales o a algún grupo de interés?**	3. ¿Cuáles son las diferencias estimadas en costos **y resultados entre la tecnología, y entre sus** comparadores?
4. ¿Cuántos pacientes podrían enfrentar perjuicios para que la **tecnología obtenga beneficio en un paciente?**	4. ¿Cuáles son las incertidumbres en torno a los costos y **las evaluaciones económicas de la tecnología y de sus** comparadores?
5. ¿Cuál es el alcance de estos beneficios y de estos **perjuicios?**	5. ¿Hasta qué punto se pueden explicar las diferencias **en los costos, los resultados o la relación costo-**eficacia por variaciones entre los subgrupos utilizando la tecnología y sus comparadores?
6. ¿La implementación o el uso de la tecnología afectan la **dignidad humana?**	6. ¿Hasta qué punto pueden ser consideradas las **estimaciones de los costos, de los resultados o de las** evaluaciones económicas?
7. ¿La implementación o el uso de la tecnología afectan la **moral del usuario, la integridad religiosa o la integridad** cultural?	
8. ¿Cómo afectan la implementación o el retiro de la **tecnología la distribución de los recursos de atención de** salud?	
9. ¿Hay factores que podrían impedir que un grupo o grupos **de personas se beneficien de la tecnología?**	

Fuente: HTA Core *Model Handbook*[19].

Tabla 25.1d Preguntas que la ET debe responder sobre cada aspecto que involucra la toma de decisiones en salud

Respecto a la prestación y la entrega de servicios de salud	Respecto a los aspectos sociales
1. ¿Cómo afecta la tecnología el proceso de trabajo actual?	1. ¿Qué tipo de cambios esperan los pacientes o los **ciudadanos con el uso de la tecnología?**
2. ¿La tecnología requiere formación del personal de salud **para su implementación?**	2. ¿Qué tipo de recursos y de apoyo son necesarios para **el paciente o el ciudadano a medida que se introduce** la tecnología?
3. ¿Qué tipo de actividades de cooperación y de **comunicación son necesarias?**	3. ¿Qué tipo de reacciones y de consecuencias puede **causar la introducción de la tecnología, en términos** generales, en la sociedad (grupos religiosos, pacientes específicos, organizaciones ciudadanas y asociaciones u otro grupo de interés)?
4. ¿Cómo están organizados el aseguramiento de la calidad y **la vigilancia del sistema de la nueva tecnología?**	4. ¿Qué tipo de influencias sociales llevan a los pacientes **a aceptar o no la tecnología?**
5. ¿Cuáles son los procesos que garantizan el acceso de los **pacientes o de los participantes a la nueva tecnología?**	5. ¿Están definidos los grupos especiales (sociales) y **de riesgo (étnico, etario, entre otros)? ¿Sus posibles** reacciones están evaluadas?
6. ¿Qué problemas de gestión, y qué oportunidades, están **relacionados con la tecnología?**	
7. ¿Qué tipo de recursos y de apoyo organizacional son **necesarios a medida que se introduce la tecnología?**	

Fuente: HTA Core *Model Handbook*[19].

Tabla 25.2 Aspectos que involucra la pregunta de política

¿Por qué se necesita la ET en este momento?

Ejemplos: Una tecnología nueva; cambios en una tecnología antigua; nuevas indicaciones de la tecnología; nuevos hallazgos; cambios estructurales o de la organización; preocupaciones respecto a su seguridad o a sus aspectos éticos o económicos.

¿Cuáles decisiones va a soportar la ET?

Ejemplos: Decisiones de inversión, de licencia de mercadeo, de inclusión o exclusión de planes de beneficios, de planeación de capacidades, de guías de buena práctica o de inversión en futura investigación.

¿Quién representa la audiencia principal del reporte de la ET?

Ejemplo: Tomadores de decisiones de políticas públicas; tercero pagador; administradores hospitalarios, médicos, pacientes o ciudadanos.

Fuente: Modificado de Busse[20].

Efectividad clínica

Determinar qué tan bien actúa una tecnología para mejorar la salud o determinar cuál es su beneficio (basándose en cambios en uno o más desenlaces o resultados finales) es parte fundamental de toda ET. El beneficio se puede abordar desde la *eficacia* y la *efectividad* de la tecnología. La eficacia se refiere al beneficio obtenido al usar una tecnología para un problema particular en condiciones ideales[14]. La efectividad, por su parte, se refiere a la capacidad demostrada de una tecnología para tener un efecto benéfico o modificar favorablemente la condición adversa en salud para la cual ha sido diseñada, en condiciones generales de rutina como la práctica diaria[14].

La escogencia y la definición de los desenlaces o de los resultados prioritarios juegan un papel primordial en la determinación de la efectividad de una tecnología. Idealmente, se debe evaluar la efectividad de la tecnología por su efecto sobre desenlaces relevantes tales como la curación completa, la supervivencia libre de progresión y el tiempo para el evento (por ejemplo, siguiente estado de enfermedad o recaída); así mismo, reportar los cambios en la gravedad, la frecuencia y la recurrencia de síntomas y los hallazgos[3,14]. Estos efectos deben presentarse en términos absolutos y relativos al comparador o la mejor tecnología disponible para la condición. En las ET de diagnóstico (pruebas o exámenes diagnósticos), la precisión de la prueba, así como los cambios benéficos en el manejo del paciente

que esta conlleva, son considerados ***desenlaces de efectividad indirecta***[19].

Cuanto mayor sea el beneficio, más probable será que la tecnología deba ser adoptada.

Efectividad clínica y seguridad de la tecnología

Determinar la seguridad de la tecnología constituye otro aspecto fundamental de toda **ET**; a mayor riesgo de efectos indeseables es menos probable que la tecnología deba ser recomendada. Muy probablemente, las tecnologías que son efectivas para una condición específica no son inocuas para los demás órganos y sistemas del organismo, y pueden causar efectos secundarios indeseables o adversos en quienes las reciben, en sus familiares o en el medio ambiente. Determinar la frecuencia y la gravedad de estos efectos potenciales asociados a la tecnología define su seguridad[3,19].

Al igual que al determinar la efectividad, aquí la escogencia y la definición de los desenlaces o de los resultados prioritarios juega un papel primordial. La seguridad se debe medir con base en eventos que suponen una amenaza para la vida o el funcionamiento del paciente, y en condiciones ideales, partir de los daños más graves o comunes hacia los de menores consecuencias.

A mayor riesgo de efectos indeseables, es menos probable que la tecnología deba ser adoptada.

Impacto económico (consumo de recursos y costo-efectividad de la tecnología)

La definición de los recursos de salud necesarios para aplicar la nueva tecnología es un paso fundamental en el desarrollo de toda **ET**. Este dominio tiene como objetivo proveer información complementaria que permita identificar si la nueva tecnología, además de competir en efectividad y seguridad con las tecnologías o las prácticas existentes, representa un mayor o un menor consumo de recursos y costos en salud. Se busca definir si la tecnología es atractiva desde el punto de vista socioeconómico: por ejemplo, si mejora la eficiencia del sistema en términos del beneficio obtenido por unidad monetaria invertida, en comparación con otras tecnologías existentes o de uso común[3,5,19].

Las nuevas tecnologías pueden ser innovadoras por su ganancia en efectividad en relación con las tecnologías previamente existentes para la misma condición; sin embargo, el consumo de recursos de salud asociado a la nueva tecnología puede ser mayor. Cuando se prevé que la nueva tecnología puede aumentar el consumo de recursos (y, por tanto, los costos asociados a su administración o su implementación), es necesario incluir en la **ET una evaluación más detallada** de la relación de los costos y de la efectividad que se obtienen con el uso de la tecnología en evaluación, así como de la diferencia en costos y efectividad que se pueden esperar sobre las tecnologías y las prácticas existentes.

Los análisis de costo-efectividad o de costo-utilidad en las **ET** tienen como propósito "identificar, medir, valorar y comparar los costos y las consecuencias de las alternativas que se consideran". En este contexto, las ***consecuencias*** **son más a menudo los resultados** de salud de las alternativas que se comparan. Estos análisis deben quedar reservados para las tecnologías que muestran ser efectivas y tienen un potencial impacto en los costos de atención, ya sea por su "ahorro" en costos a largo plazo o porque prometen mayor "eficiencia" (mayor beneficio obtenido por peso invertido) que las prácticas actuales u otras tecnologías usadas para la misma condición, no solo por su costo de adquisición o de aplicación inicial[5,19]. **La** definición de la costo-efectividad de una tecnología y sus alternativas se establece de acuerdo con el umbral de disposición a pagar que define el Estado para adoptar una tecnología dentro de su sistema de salud.

Los elementos que hacen parte de los análisis económicos de costo-efectividad y costo-utilidad se describen en el **capítulo 23 "Introducción a la** evaluación económica de tecnologías en salud" presente libro, y son aplicables a los contenidos de este dominio.

Si la tecnología mejora la eficiencia en términos de beneficio obtenido por unidad monetaria invertida, en comparación con otras tecnologías existentes, será más atractiva para el sistema de salud.

Aspectos éticos relacionados con la adopción de la tecnología (balance de beneficios y riesgos; impacto sobre la equidad)

Este dominio busca abordar aspectos relativos a los aspectos éticos que pueden surgir como resultado del uso de la tecnología en evaluación, y que ameritan una consideración especial al momento de la toma de decisiones.

Un problema de salud puede afectar desproporcionalmente a algunos grupos de la sociedad; los beneficios, los daños y los costos de las tecnologías disponibles para solucionar un problema pueden variar a través de los grupos; así mismo, las consideraciones sobre la implementación de nuevas tecnologías pueden variar de un grupo a otro. Algunos grupos de la sociedad, como los que viven en áreas rurales y remotas, o las personas de bajo nivel socioeconómico, suelen ser los menos favorecidos con la adopción de tecnologías que requieren altos niveles de complejidad para su implementación y su funcionamiento; por tanto, merecen una consideración seria al momento de la toma de decisiones, para evitar la inequidad.

Por otra parte, la implementación de algunas tecnologías en salud trae consigo otros cuestionamientos éticos por la pérdida de la naturalidad y la manipulación artificial que involucran, tales como las relacionadas con pruebas genéticas, el uso de células madre para la regeneración de huesos y tejidos o los sistemas para soporte de la vida en pacientes críticamente enfermos, entre otras[5].

Con esta perspectiva, hay dos tipos de tecnologías: 1) unas que se pueden llamar ***tecnologías sin controversias éticas*** y 2) otras ***éticamente controversiales***[19]. En el primer caso, en la ET los aspectos éticos pueden ser abordados con la definición específica de preguntas o estamentos en los consentimientos informados, por ejemplo. En el segundo caso, se requiere un análisis ético más complejo, que se basa en los principios éticos de respeto por la autonomía (respeto por las decisiones del individuo respecto a su propia vida), beneficencia (hacer el bien y fomentar una buena vida y buscar el equilibrio de los posibles beneficios de una acción contra sus posibles efectos nocivos), no maleficencia (obligación de evitar causar daño intencionalmente) y justicia (asegurar una distribución equitativa de todos las beneficios básicos y de las cargas)[5].

La comprensión ética puede ser objeto de estudios empíricos: por ejemplo, encuestas o investigación cualitativa respecto a la apreciación que tiene el paciente sobre los efectos asociados a la tecnología, no solo en la salud, sino en su moral y sus aspectos socioculturales. Por lo tanto, es obvio que en las ET se debe considerar la ética en una dimensión interdisciplinaria que relacione los distintos elementos de la tecnología objeto de evaluación.



La adopción de una tecnología es inaceptable si conlleva un desacuerdo moral en relación con principios éticos tales como la autonomía, la no maleficencia, la beneficencia o la justicia.

Aspectos sociales (valores y preferencias de los pacientes y aceptabilidad de los implicados)

Este dominio busca brindar información sobre cómo la sociedad percibe la tecnología. Se centra en las consideraciones, las preocupaciones y las experiencias de los pacientes y sus personas cercanas, antes, durante y después de la aplicación de la tecnología. En él se describe cómo la tecnología se moldea en diversos ámbitos sociales en los que los pacientes la utilizan (hospitales, médicos generales, de vida cotidiana, hogares, escuelas y lugares de trabajo), y qué significados específicos da la gente a la tecnología[19].

Es probable que las principales partes interesadas de la sociedad encuentren aceptable la adopción de la tecnología, dada la importancia relativa que les asignan a las consecuencias deseables e indeseables de la tecnología, la temporalidad de los beneficios, los daños y los costos, y sus valores morales. La aceptabilidad puede reflejar quiénes se benefician (o son perjudicados) y quiénes pagan (o ahorran); y cuando ocurren los beneficios, los efectos adversos y los costos. La no aceptación puede ocurrir debido a que algunas de las partes interesadas no están de acuerdo con la distribución de los beneficios, los daños y los costos, o no aceptan los costos o los efectos indeseables a corto plazo por los efectos deseables (beneficios) en el futuro, o les asignan más valor (***importancia relativa***) a las consecuencias indeseables, en comparación con las consecuencias deseadas o los costos de una opción[5,19].

Si una tecnología poco aceptada es adoptada por el sistema, es necesario incluir una estrategia de implementación que aborde las dudas de los interesados.

Mientras menos aceptable sea la adopción de la tecnología para las principales partes interesadas, es menos probable que esta sea recomendada.

Aspectos organizacionales

Comprender los aspectos organizativos relacionados con la adopción de la tecnología puede revelar desafíos esenciales y los obstáculos en la implementación de tecnologías de la salud. Este dominio busca mostrarle al tomador de decisiones los efectos que sobre la organización puede llegar a tener el hecho de adoptar la tecnología. Se centra en el análisis de los modelos de prestación de la tecnología, de los procesos, los recursos, la gestión y las cuestiones culturales, en los planos intraorganizacional e interorganizacional y del sistema de atención de salud. Este aspecto debe tener un enfoque analítico e ilustrar diferentes dimensiones del proceso en la interacción entre la tecnología de la salud y la organización.

Las organizaciones del sistema de atención de salud son sistemas sociales donde las personas actúan y reaccionan; las percepciones, el análisis de los objetivos, los recursos y los tipos de organizaciones del sistema de atención de salud se hallan en constante movimiento y son continuamente objeto de negociaciones en curso. Por su parte, los proveedores de salud trabajan, cooperan y compiten entre sí, y construyen sistemas complejos en rutinas y renovaciones[19]. Lograr el control de los efectos de diversas acciones organizacionales en la adopción de una tecnología es una tarea compleja y ambiciosa; por tanto, la evaluación de los aspectos organizacionales en las ET debe centrarse en los objetivos más relevantes, para que al adoptar la tecnología se logren los efectos deseados[5,19].

Si la tecnología implica cambios organizacionales sustanciales, es menos posible lograr las condiciones de implementación necesarias para que esta sea efectiva.

Aspectos legales

En este dominio se consideran los acuerdos implícitos y explícitos del fabricante (o el vendedor) y el comprador de la tecnología y los requisitos legales, como la autorización, la garantía y la regulación del mercado. Incluye también preguntas sobre los derechos básicos de los pacientes, tales como la autonomía, el consentimiento informado, la privacidad y la confidencialidad. Para definir tales aspectos, se escrutan en la legislación las fuentes jurídicas pertinentes y los convenios nacionales o internacionales en la materia[19].

PROCESO DE LA EVALUACIÓN DE TECNOLOGÍAS

El contexto en el que se hace la ET determina los métodos que se deben usar, así como la extensión y la amplitud de la evaluación. El alcance y el nivel de detalle de las ET varían considerablemente dependiendo de quién ha comisionado la evaluación y por qué (por ejemplo, hospitales, para definir la inversión en la tecnología; o el sistema de salud, para definir planes de cobertura de servicios de salud). No siempre se requiere incluir todos los dominios presentados anteriormente con el mismo nivel de detalle: algunos efectos pueden ser ya conocidos o pueden no ser motivo de preocupación[17]. Es, entonces, crucial explicar el contexto donde se da la ET para que el usuario del reporte (lector) pueda entender sus contenidos, así como definir la utilidad de este para sus necesidades de información. La información del contexto también ayuda a definir la pregunta de política que suscita la ET, y, por tanto, es un aspecto clave del reporte de la ET[20].

En el curso de una ET se incluyen muchos componentes; algunos son inherentes al proceso, mientras que otros se refieren a los análisis. A menudo están interconectados y pueden ser manejados en paralelo, de manera solapada o relacionados con los pasos anteriores. Sin embargo, una vez se ha identificado la necesidad de la ET y esta se ha comisionado a sus desarrolladores, se pueden identificar cinco fases en el proceso[19]. La **figura 25.1** muestra un esquema que resume los pasos que comprende cada fase.

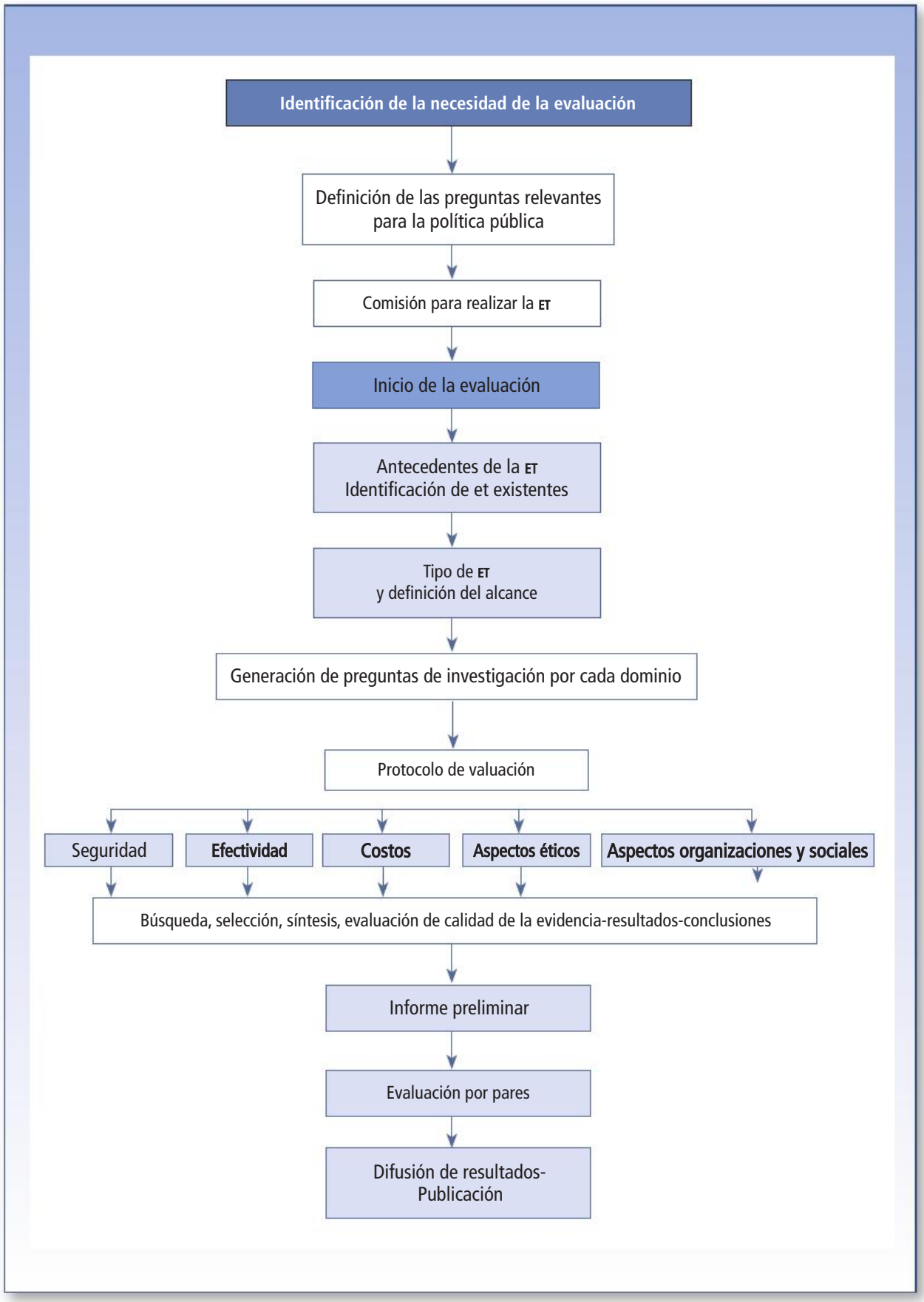


Figura 25.1 Proceso en la ET.
Fuente: Modificada de Busse[20].

Fase 1. Definición del proyecto y de sus participantes

Esta fase se lleva a cabo en estrecha comunicación entre el grupo evaluador y quien ha comisionado la ET. Parte de la definición de la pregunta de política (si esta no se ha definido previamente como parte de la priorización o la comisión de la ET), lo que va estrechamente ligado a definir el tipo de evaluación que se quiere hacer (por ejemplo, orientada a la tecnología) y el tipo de informe que se espera producir (por ejemplo: amplio, enfocado, rápido, completo). La **tabla 25.2** presenta los aspectos que involucran la definición de la pregunta política de una ET[20].

El paso siguiente consiste en obtener información sobre los antecedentes de la tecnología que permita determinar el alcance de la ET, y decidir los aspectos del problema que es más relevante evaluar en cada dominio de la evaluación[19,20]. El alcance de la ET se estructura a partir de la pregunta que se busca responder[17,19,20], por lo cual es práctico definirlo en términos de población, intervención (tecnología), comparadores y desenlaces para evaluar, así:

- Población objeto de la tecnología: Descrita en términos de su condición de salud (condición para la cual se busca usar la tecnología), con énfasis en los posibles subgrupos de interés o hacia los que se deben tener consideraciones especiales al usar la tecnología.
- La intervención: Características técnicas de la tecnología a partir de lo que se conoce de ella en cuanto a su mecanismo de acción, su dosis, su modo de uso, su estado de desarrollo, sus ventajas potenciales respecto a otros medicamentos de su clase (comparadores), etc. Aquí, la información que proporcione el fabricante puede ser de gran ayuda.
- Comparadores: Otras tecnologías disponibles para la condición de salud objeto de la tecnología, enfocándose en la mejor alternativa disponible, y mostrando sus ventajas y sus desventajas conocidas hasta el momento.
- Resultados: En términos de los desenlaces de salud de los cuales se espera que la tecnología modifique, priorizados por su importancia para el paciente.

La definición del tipo de evaluación, el alcance y el tipo de informe que se quiere producir llevará al paso que se puede considerar el último de esta fase: determinar los participantes o el grupo desarrollador que requiere el proyecto, como: expertos metodológicos (por ejemplo, un epidemiólogo clínico), experto temático (por ejemplo, alguien de reconocida experiencia en el área clínica de la condición objetivo de la tecnología en evaluación), experto en búsquedas de evidencia (por ejemplo, un bibliotecario o su equivalente), experto en evaluación económica (por ejemplo, un economista de salud), experto en ética y equidad (por ejemplo, un sociólogo o un bioeticista), entre otros. Varios de ellos actuarán como consultores para determinados aspectos de la evaluación, mientras que otros tendrán la función de desarrollar (por ejemplo, liderar, ejecutar) la evaluación y permanecerán durante todo el desarrollo del proyecto.

Fase 2. Establecimiento (diseño) del protocolo

Esta fase pretende, a partir del establecimiento de las preguntas que se van a resolver, marcar los lineamientos que guiarán el proceso para encontrar las respuestas a esas preguntas, que serán el resultado de la ET. Aquí no se busca seguir un protocolo de investigación detallado[19,20]. Esta fase comprende cuatro pasos que llevan a definir el protocolo del proyecto:

Formulación de las preguntas de investigación

Formular las preguntas de investigación significa especificar cuál es la pregunta en política de salud, en términos de efectividad, seguridad, y de aspectos éticos, organizacionales y económicos. Por tanto, la formulación de dichas preguntas está íntimamente ligada con la información obtenida como parte de los antecedentes. Las preguntas deben especificar la población blanco de la tecnología en términos de su condición, y, especialmente, definir los desenlaces de interés para la evaluación, los cuales pueden diferir según el aspecto evaluado, pero siempre deberán centrarse en el paciente (por ejemplo, el impacto de la tecnología en los diferentes aspectos de su calidad de vida)[21].

La seguridad y la efectividad de una intervención serán medidas en términos de los beneficios en salud para el paciente (por ejemplo, reducción de la morbilidad, impacto en la mortalidad, calidad de vida, discapacidad,

etc.), mientras que los aspectos éticos y sociales pueden ser medidos en términos de satisfacción, aceptación o preferencias del paciente; los efectos sobre la organización pueden evaluarse a partir de resultados relativos al sistema (por ejemplo, duración de la estancia hospitalaria, o infraestructura requerida), y, por último, los aspectos económicos de mayor interés se basarán en el consumo de recursos, y sus costos, en relación con los desenlaces (por ejemplo, costo-efectividad y costo-utilidad; y razón incremental de costo-efectividad, en comparación con otras alternativas disponibles para la condición)[5].

El establecimiento claro de las preguntas guiará los contenidos de la evaluación (cuáles dominios serán evaluados y cuáles no), así como la definición de los criterios para la búsqueda de información y las fuentes que se deben consultar. Por tanto, las preguntas deben cumplir con las características de una pregunta bien formulada (por ejemplo, que sea factible responderla, que sea clara y específica y que contenga los elementos de la estructura ***población de interés, intervención, comparadores, resultados y tiempo*** (estructura PICOT, por las siglas en inglés de ***population of interest, intervention,*** *comparators, outcomes and time*). El capítulo 2 de este libro , "La pregunta de investigación", presenta en detalle las características de una buena pregunta.

Metodología

Es primordial, para garantizar la calidad y la transparencia del proceso, definir la metodología que se seguirá para obtener la información que dará respuesta a las preguntas de investigación formuladas en cada uno de los aspectos (dominios) que incluirá la ET.

Aunque los métodos pueden diferir según el aspecto en evaluación, existen algunos principios metodológicos comunes, que se pueden resumir en cuatro pasos: 1) definir las fuentes de información, 2) la búsqueda, la selección y la evaluación de la información, 3) la síntesis de los datos obtenidos y 4) la valoración de la calidad (***certeza***) del cuerpo total de evidencia.

Fuentes de información

Para los diferentes dominios de la evaluación pueden ser útiles distintas fuentes de datos. Con el fin de darles una solución a los problemas o los cuestionamientos planteados, los evaluadores deben tratar, en lo posible, de usar datos provenientes de la literatura científica (***resultados de investigación primaria***); cuando ello no es posible, el uso de bases de datos, registros de rutina, e, incluso, la investigación propia basada en entrevistas, encuestas u observación puede ser apropiada[19]. Otra fuente de información es la que proporciona el fabricante de la tecnología (por ejemplo, folletos de historia del medicamento, o estudios clínicos y reportes de seguridad)[17,19,20,22].

Los tipos de estudios que soportan las decisiones también varían de dominio a dominio; el estándar de oro para evaluar la efectividad de una tecnología en salud es la investigación basada en experimentos clínicos aleatorizados (ECA), pero las inferencias sobre la efectividad de las tecnologías diagnósticas deberán considerar si ligan los estudios de precisión de la prueba con los de efectividad del tratamiento, pues los ECA que evalúan la prueba y el tratamiento son muy escasos en la literatura. Así mismo, los daños asociados al uso de una tecnología no siempre están bien reportados en ECA; la información sobre los eventos adversos serios a largo plazo son más comúnmente reportados en estudios de seguimiento (estudios de fase IV, o posmercadeo) a largo plazo, como los estudios de cohortes, o ser tan poco comunes que solo se han documentado en series de casos. Para otros dominios, como los relacionados con aspectos sociales y organizacionales, los estudios cualitativos son la principal fuente de evidencia, así como los estudios observacionales de diferentes diseños pueden también proveer información de relevancia.

Con independencia de la fuente de información que se use, se deben seguir una aproximación sistemática y la estrategia seguida para su identificación; y la manera como se obtiene la información debe ser documentada. Si se usa literatura científica, se deben seguir principios sistemáticos para su búsqueda, tales como los desarrollados por la Colaboración Cochrane para identificar evidencia[23], y los cuales pueden ser aplicados para todos los aspectos de la ET, y no solo para la efectividad y la seguridad. Es necesario, entonces, desarrollar una estrategia de búsqueda basada en las preguntas que se quiere responder, las cuales guiarán la definición de los criterios de selección de los estudios.

Las estrategias de búsqueda, así como las bases de datos, van a diferir según el dominio que se esté evaluando, y, por tanto, estas deben

documentarse de forma independiente. Documentar las fuentes de información es un aspecto crucial para la trasparencia de la ET, por lo cual se debe incluir en la documentación: cuáles fuentes se consultaron, qué periodo se cubrió en la búsqueda, cómo se hizo la búsqueda (estrategias, palabras clave o términos de búsqueda) y cuándo fue realizada la búsqueda (fecha)[20,23]. Mayores detalles sobre cómo llevar a cabo una búsqueda de literatura sistemática y eficiente se presentan en el **capítulo 3** "Búsqueda de literatura médica", de este libro.

Selección y evaluación de la información

Los pasos por seguir a la hora de seleccionar la literatura que finalmente se incluirá para responder las preguntas de la ET son parte de los métodos que deben estar claramente definidos en el protocolo. Los criterios de selección, así como quién realizará el proceso, se deben describir aquí, y documentarse, para garantizar la transparencia y la reproducibilidad del proceso. Los criterios de selección serán definidos, entre otros aspectos, por el tipo de estudios, el tipo de población y el tipo de intervenciones evaluadas. Dichos criterios, al igual que el proceso de selección, deben estar definidos para todo tipo de evidencia que se incluya en la ET, y no solo para la que se incluye en los dominios de efectividad y seguridad de la tecnología[20].

Por otra parte, una vez seleccionados los estudios, se requiere apreciarlos críticamente, para definir su calidad metodológica y la validez de sus resultados. Solo los estudios que cumplan estándares de buena calidad deberían soportar las decisiones sobre uso de las tecnologías en salud; sin embargo, puede haber excepciones, y los evaluadores deben aclarar el uso que harán del resultado de la apreciación crítica y si este definirá o no la inclusión de los estudios en la ET.

El protocolo debe definir quién o quienes harán la evaluación (por ejemplo, dos revisores independientes) y qué criterios e instrumentos de los disponibles en la literatura biomédica se usarán para llevar a cabo la apreciación crítica de los diferentes tipos de estudios (por ejemplo: herramienta Cochrane para evaluación de riesgo de sesgo en ECA[24]; AMSTAR, para evaluar revisiones sistemáticas de literatura[25]; QUADAS, para estudios de pruebas diagnósticas[26], entre otros). La identificación del instrumento más apropiado para apreciar críticamente los estudios cualitativos es un reto que enfrentan los evaluadores en este punto, ya que la evidencia del efecto de la tecnología sobre los aspectos sociales y éticos proviene, en su mayor parte, de estudios de investigación cualitativa, y es necesario usar el instrumento adecuado para los tipos de estudios que se van a evaluar[27].

Los elementos del proceso de selección de estudios y apreciación crítica de la evidencia se describen en detalle en los **capítulos 20** "Revisiones sistemáticas y metaanálisis de la literatura" y **35** "Apreciación crítica de la literatura", de este libro.

Síntesis de la evidencia

El siguiente paso en el protocolo es definir cómo se hará la síntesis de la información. Este proceso es crucial para la validez de la información que se presente sobre cada dominio. Es pertinente presentar una descripción de la forma como se propone resumir y estimar el efecto que el uso de la tecnología tiene sobre los desenlaces y los resultados de interés seleccionados, de forma ***cualitativa*** (describiendo los hallazgos en forma narrativa) o de forma ***cuantitativa*** (haciendo análisis estadísticos).

Es necesario partir de información válida y de datos apropiados para cada desenlace, por lo que es recomendable que la extracción de datos sea hecha por dos evaluadores, de forma independiente (lo que dará validez y trasparencia al proceso), y que se cuente con formatos prediseñados para incluir la información relevante para cada pregunta en evaluación (por ejemplo, la población estudiada, las intervenciones que se comparan, los resultados evaluados y la medida del efecto sobre los desenlaces de interés), al igual que información general relacionada con el estudio (referencia completa, lugar de realización, patrocinio, aspectos éticos, etc.)[20].

Es probable que como resultado de la búsqueda se identifiquen revisiones sistemáticas de la literatura (RSL) que respondan la pregunta de interés y hayan evaluado los desenlaces seleccionados como relevantes, y que, además, estén vigentes (con fecha de búsqueda no mayor de dos años que la fecha de evaluación). En este caso, la información que aporta la RSL más reciente y completa puede ser tomada en cuenta para basar la evaluación, y las estimaciones del efecto del uso de la tecnología que dicha RSL reporte, ser usadas directamente.

También es probable que las RSL halladas cumplan las características de la pregunta y los desenlaces de interés, pero se encuentren desactualizadas; en tal caso, los evaluadores deben esforzarse por identificar nuevos estudios elegibles para ser incluidos en la RSL, y si los encuentran, actualizar la síntesis de la evidencia incorporando los resultados de nuevos estudios. En casos de síntesis metaanalítica, será necesario evaluar la pertinencia de incluir o no los resultados de los nuevos estudios (por ejemplo, si se ha descartado la heterogeneidad clínica y se puede aceptar la heterogeneidad estadística en los resultados). También es posible que no se encuentren en la literatura revisiones sistemáticas que hayan respondido la pregunta de interés, pero sí, estudios primarios (por ejemplo, ECA, cohortes) que aporten información sobre el desenlace de interés en la evaluación; entonces, será pertinente evaluar la posibilidad de hacer síntesis cuantitativa de sus resultados (por ejemplo, metaanálisis) para estimar una medida de efecto agrupada basada en los resultados que dichos estudios reportan.

Los métodos para estimar el efecto de la intervención sobre cada desenlace a partir del metaanálisis han sido ampliamente estudiados y desarrollados por varios autores, y es apropiado seguir los lineamientos propuestos por la Colaboración Cochrane para tal fin[23].

La definición clara de los criterios y de los procedimientos para llevar a cabo un metaanálisis hace parte del **capítulo 20** "Revisiones sistemáticas y metaanálisis de la literatura", del presente libro. Las medidas de efecto agrupado comúnmente estimadas son el riesgo relativo (RR) y el riesgo relativo indirecto (OR, por las siglas en inglés de *odds ratio*) para un resultado dicótomo (por ejemplo, mejoría-no mejoría), la diferencia de medias estandarizada (SDM, por las siglas en inglés de *standardized mean difference*) para una variable de resultado continuo (por ejemplo, colesterol en mg/dl) o *hazard ratio* (HR) para los resultados relacionados de tiempo evento (conocido como *análisis de sobrevida*).

Para definir el impacto que puede llegar a tener la intervención (o el uso de la tecnología), basado en estudios individuales, las medidas más útiles son la reducción absoluta del riesgo (RAR) y el número necesario para tratar (NNT). La estimación y la interpretación de dichas medidas se explican detalladamente en el **capítulo 9** "Medidas de frecuencia, de asociación y de impacto", del presente libro.

En esta sección del protocolo, también es importante definir cómo se presentarán los datos que soportan las conclusiones de la evaluación (por ejemplo: tablas de evidencia, tablas de resumen de datos, etc.). La idea es presentar de forma estructurada la información más relevante respecto a la validez, la calidad y los resultados de los estudios que se han usado en la evaluación; los elementos que se vayan a incluir dependerán del tipo de estudio y del alcance de la evaluación.

Evaluación del cuerpo global de evidencia

La calidad de la evidencia hallada sobre el efecto de la tecnología tiene implicaciones muy importantes para la toma de decisiones, por lo que la evaluación del cuerpo global de la evidencia es un paso obligatorio en las ET. Los grupos que hacen ET deben esforzarse en describir *a priori* cómo se evaluará la calidad de la evidencia que se encuentre sobre la tecnología en evaluación y el uso que darán al resultado de dicha evaluación. El grupo de estudios seleccionados como la "mejor evidencia disponible" para resolver las preguntas planteadas en la ET se conoce como *cuerpo de evidencia*[17].

El cuerpo de evidencia se caracteriza por la combinación de factores que se han discutido previamente, tales como el tipo de diseño, la calidad metodológica y la dirección del efecto. Adicionalmente, otros factores, como el número de estudios, el tamaño del efecto, la homogeneidad y la consistencia de sus resultados, son importantes para determinar la calidad global de la evidencia que soporta los resultados de la evaluación[17]. El reto radica en juzgar el cuerpo de evidencia a partir de todos esos factores en los estudios individuales.

La aproximación propuesta por el grupo de trabajo "Grading of Recommendations Assessment, Development and Evaluation (GRADE work group)"[28] es una de varias que se han descrito para estandarizar la forma como los investigadores y los generadores de síntesis de evidencia para ET o guías de práctica clínica (GPC) evalúan la calidad de la evidencia que soporta las decisiones clínicas o de política pública. El GRADE WG propone evaluar todas estas características del cuerpo de evidencia para determinar "la certeza de la evidencia"; o sea, la probabilidad de que el efecto sea sustancialmente diferente de lo hallado en la investigación[29].

Esta aproximación es actualmente la más usada en organizaciones dedicadas al desarrollo de GPC, como, por ejemplo: National Institute of Clinical Evidence (NICE), la Organización Mundial de la Salud (OMS, o WHO, por las siglas en inglés de World Health Organization), Scottish Intercollegiate Guidelines Network (SIGN), y más recientemente, recomendada por las asociaciones de agencias de evaluación de tecnologías como EUNETHTA, entre otras.

La aproximación GRADE es una forma de evaluar la calidad de la evidencia en la ET que soporta las decisiones clínicas o de política pública.

La aproximación (GRADE) sugiere escoger entre los cuatro o cinco desenlaces más relevantes para la toma de decisiones (esto debe coincidir con lo que los expertos y quienes comisionan la ET han definido al generar las preguntas PICOT que responderá la evaluación) sobre los cuales valorar la calidad del cuerpo de evidencia, y definir así la "certeza en la estimación del efecto" encontrado (valoración de la probabilidad de que el efecto sea sustancialmente diferente de lo que se halló en la investigación). El GRADE considera para los desenlaces de efectividad de las intervenciones, por ejemplo, que los ensayos clínicos aleatorizados proveen la evidencia de más alta calidad; sin embargo, la certeza en el cuerpo de evidencia que proviene de los ECA puede verse afectada por cinco aspectos:

a) Limitaciones en el diseño que los lleven a un alto riesgo de sesgo.
b) Inconsistencia en los resultados de los estudios (diferencias que los hacen realmente heterogéneos).
c) Cuando los estudios proporcionen evidencia indirecta (la pregunta que se aborda en la ET es diferente de la evidencia disponible en cuanto a la población, la intervención, el comparador o los desenlaces evaluados).
d) Imprecisión en los resultados (los estudios incluyen relativamente a pocos pacientes y pocos eventos, los resultados son imprecisos, y, por lo tanto, tienen intervalos de confianza amplios en torno al estimador del efecto).
e) Sesgo de publicación (la estimación del efecto está subestimada o sobreestimada de forma sistemática, y se debe a la publicación selectiva de estudios). Ante la presencia de alguno de ellos (o de varios), la "certeza" en la evidencia para el desenlace en evaluación se reduce[29]. La explicación detallada sobre cómo juzgar cada uno de los mencionados aspectos hace parte del capítulo de RSL.

Al final de la evaluación del cuerpo de la evidencia para cada desenlace, se tendrá el resultado sobre la certeza en las estimaciones, la cual puede ser "alta", "moderada", "baja" o "muy baja". La **tabla 25.3** resume el signi-

Tabla 25.3 Calificaciones de la certeza en la evidencia según la aproximación GRADE

Calificación	Definiciones	Implicaciones
Alta	La investigación provee un muy buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente distinto es muy baja.	La evidencia es una excelente base para la toma de decisiones sobre la implementación o no de la intervención (uso de la tecnología). Es probable que no se requieran evaluación de impacto ni monitoreo del impacto si la intervención es implementada.
Moderada	La investigación provee un buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente distinto es moderada.	La evidencia es una buena base para la toma de decisiones sobre la implementación o no de una intervención. Es probable que se requiera monitoreo del impacto, y podría justificarse una evaluación del impacto, si la intervención es implementada.
Baja	La investigación provee alguna indicación del efecto probable; sin embargo, la probabilidad de que el efecto sea sustancialmente distinto es alta.	La evidencia provee alguna base para la toma de decisiones sobre la implementación o no de una intervención. Sin embargo, se justifica hacer una evaluación del impacto si la intervención es implementada.
Muy baja	La investigación no provee una indicación confiable sobre el efecto probable. La probabilidad de que el efecto sea sustancialmente distinto es muy alta.	La evidencia no es una buena base para la toma de decisiones sobre la implementación o no de una intervención. Preferiblemente, la intervención no debe ser implementada; si lo es, se requiere evaluar su impacto.

ficado de cada uno de dichos resultados y sus implicaciones en la toma de decisiones[29].

Para presentar la síntesis de la evidencia encontrada y la valoración que se hace de su calidad, el GRADE WG propone hacer una tabla de resumen de hallazgos (SoFt, por sus siglas en inglés), la cual se puede crear a partir de la información obtenida usando el programa GRADE pro (disponible en: http://tech.cochrane.org/revman/other-resources/gradepro/download). Es pertinente que el grupo evaluador desde el protocolo defina la necesidad de hacer una tabla por cada subgrupo de interés (por población o por intervención) que se incluya en la ET, y qué desenlaces serán evaluados para el subgrupo en cuestión (no más de cinco). Para completar correctamente estas tablas, se puede consultar en el *Cochrane Handbook for Systematic Reviews of Interventions*[30].

Si bien hasta aquí se han presentado los lineamientos metodológicos generales que son comunes a todos los aspectos que evalúa una ET, el protocolo del proyecto deberá dar cuenta, específicamente, de los métodos que se seguirán para obtener respuestas a las preguntas abordadas en cada dominio en particular (por ejemplo: efectividad, seguridad, aspectos éticos, sociales y organizacionales), según del tipo de informe que se esté preparando (ET completa o ET breve).

Si la evaluación es completa y se requiere un análisis económico de costo-efectividad o de costo-utilidad, es necesario describir en el protocolo los métodos que se van a seguir para su conducción. La **tabla 25.4** presenta los aspectos principales que se deben describir en una evaluación económica para ET. Los detalles metodológicos que involucran cada uno de dichos aspectos en los estudios de evaluación económica son ampliamente descritos en el **capítulo 23** "Introducción a la evaluación económica de tecnologías en salud", del presente libro.

Particularidades metodológicas para la evaluación de tecnologías de diagnóstico

Como ya se mencionó, las tecnologías diagnósticas se pueden usar en diferentes formas (por ejemplo, identificación de una enfermedad, monitoreo del progreso de una enfermedad o de su tratamiento, tamización o identificación temprana), y todo ello debe verse reflejado en la evidencia que las soporta. La evaluación de una tecnología de diagnóstico debe incorporar tanto la evidencia en la precisión de la prueba como los cambios en la salud y los costos que resultan de las decisiones tomadas sobre el tratamiento, con base en los resultados de la prueba. Por tanto, el grupo evaluador que conduce la ET debe asegurar los métodos para identificar estudios que evalúan la precisión de la prueba, así como la de la evidencia sobre el efecto del tratamiento (usualmente, el más aceptado o recomendado para la condición); o sea, ubicar los ECA o las RSL que evalúan el efecto del tratamiento. En este paso, contar con una estrecha asesoría por parte del experto clínico del grupo evaluador enriquece y facilita el proceso; él es quien podrá establecer las fuentes (por ejemplo, GPC) de información sobre las intervenciones terapéuticas comúnmente aceptadas o más recomendadas para la condición objetivo de la tecnología que se está evaluando[22].

Los principios generales que guían la evaluación de la efectividad clínica y la costo-efectividad de una prueba diagnóstica son los mismos que ya se han presentado para otras tecnologías. Sin embargo, se requiere una consideración especial de los métodos de síntesis de la evidencia (análisis), pues la evidencia de efectividad de una prueba diagnóstica, así como su costo-efectividad, debe incluir consecuencias para los que fueron incorrectamente diagnosticados como resultado de la prueba y para los que fueron correctamente diagnosticados[22,29].

Al igual que para otro tipo de tecnologías, los ECA tienen el potencial de capturar todo el curso de cuidado que envuelven las tecnologías diagnósticas, pero su factibilidad y su disponibilidad en la literatura pueden ser bastante limitadas, por lo cual es pertinente evaluar estudios que han seguido otros diseños, con el propósito de evaluar el comportamiento y la precisión de la tecnología diagnóstica de interés[22].



La ET de una tecnología de diagnóstico debe incorporar tanto la evidencia en la precisión de la prueba como los cambios en la salud y los costos que resultan de las decisiones tomadas sobre el tratamiento, con base en los resultados de la prueba.

Fase 3. De investigación y análisis de la información

En esta fase se reúnen los procesos que han sido propuestos en el protocolo para dar res-

Tabla 25.4 Aspectos que se deben describir sobre las evaluaciones económicas en la ET

1. **Marco del análisis: Descripción del tipo de análisis y su justificación. Formulación de la pregunta económica, la cual debe** coincidir con la pregunta definida en el alcance de la et, e incluir, además, el aspecto económico que se quiere evaluar (por ejemplo, la costo-efectividad relativa o la razón incremental de costo-efectividad).
2. **Perspectiva del análisis: En las et, la perspectiva deberá ser de la sociedad, pero es posible considerar el tercero** pagador.
3. **Alternativas para comparar: Justificación de la selección de las alternativas con las que se va a comparar la tecnología** en evaluación (idealmente, la más efectiva o la más eficiente conocida). Debe coincidir con los comparadores establecidos en el alcance de la et.
4. **Desenlaces: Descripción detallada de los desenlaces de efectividad que serán evaluados; si bien debe coincidir con los** desenlaces presentados en el alcance de la et, aquí debe haber una definición operativa de cada desenlace.
5. **Medidas de utilidad: Pertinencia de evaluar el impacto de la tecnología usando medidas sintéticas (por ejemplo, años** de vida ajustados por calidad [QALY, por las siglas en inglés de *quality adjusted life Years*] o años de vida ajustados por discapacidad [DALY, por las siglas en inglés de *disability adjusted life years*]). Es necesario describir con claridad cómo se obtendrán las fuentes de utilidad o las preferencias que se utilizarán para su cálculo, etc.
6. **Costos: Descripción del tipo de costos que se van a incluir (directos, indirectos), los cuales van en relación con la perspectiva** del análisis.
7. **Horizonte de tiempo: Indicar el horizonte de tiempo y su justificación. Idealmente, que sea tan largo como para capturar** todos los efectos posibles de la intervención o el uso de la tecnología en la salud del paciente; sin embargo, los horizontes de largo plazo (por ejemplo, a la expectativa de vida), deben reservarse para cuando la evidencia de efectividad es completa o permite su modelamiento a largo plazo.
8. **Descuento: Es necesario establecer una tasa de descuento si los costos y las consecuencias ocurren en diferentes momentos** del tiempo. Se debe especificar aquí la tasa de descuento usada para costos y desenlaces, con su respectiva justificación (por ejemplo, el 3 % o el 5 %).
9. **Fuente de información: Descripción de las fuentes de información para determinar el curso de enfermedad, los desenlaces,** la efectividad de la tecnología y sus comparadores. Es necesario describir las fuentes que se usarán para definir el consumo de recursos y la forma como se recolectará la información de costos (facturas, levantamiento a partir de fuentes primarias, consulta de bases de precios, tarifas, etc.).
10. **Tipo de modelamiento: Breve descripción del modelo o del análisis de decisiones que se va a realizar. Considerar su** plausibilidad y su coherencia con la enfermedad objeto de la tecnología, y buscar su comparabilidad con otras evaluaciones económicas sobre la misma tecnología y la misma indicación.
11. **Fuentes de variabilidad: Si se identifican grupos poblacionales en los cuales pueden anticiparse diferencias en los** resultados (costos, efectividad, utilidad). Describir las fuentes de variación que ameriten análisis independientes o de subgrupos.
12. **Fuentes de incertidumbre: Identificar las posibles fuentes de incertidumbre (en los parámetros de efectividad o de** consumo de recursos (costos) o en los supuestos del modelo) que ameriten una evaluación para validar la robustez de los resultados.
13. **Métodos para evaluar el efecto de la variabilidad y la incertidumbre: Describir los métodos para el análisis** de la incertidumbre (análisis de sensibilidad simple, análisis de umbral, análisis de extremos y análisis de sensibilidad probabilístico, etc.).

puesta (generar resultados) a los interrogantes planteados en la ET. Corresponde a la ejecución de la evaluación misma en la que se selecciona y se evalúa la información; síntesis de la evidencia, y evaluación del cuerpo global de la evidencia, para cada dominio. Para las ET completas, durante esta fase se recolecta (se *hace levantamiento*) de la información de costos que nutrirá el análisis económico y se llevan a cabo los análisis necesarios para responder a los interrogantes de este dominio. Aquí, el papel del epidemiólogo clínico del grupo es decisivo para interpretar resultados y definir conclusiones sobre los dominios de seguridad y efectividad de la tecnología. La aplicación de conceptos como ***dirección del efecto***, ***tamaño del efecto*** y *precisión de la estimación del efecto* es clave para determinar el beneficio y el daño que se pueden esperar con el uso de la tecnología.

Durante la misma fase, el grupo evaluador establece la necesidad de obtener datos adicionales a los hallados en la literatura, o la de adelantar investigación primaria para los aspectos en los cuales no se ha identificado evidencia. El inicio de investigaciones en el contexto de la ET se considera únicamente si el conocimiento que se ha obtenido con la revisión de la literatura prueba ser inadecuado o se considera insufi-

ciente en los dominios primordiales para tomar una decisión sobre la tecnología en cuestión[5]. Por ejemplo, se quiere evaluar si la adición de nevirapina al tratamiento con zidovudina en madres gestantes positivas para el virus de la inmunodeficiencia humana (VIH) durante el último trimestre del embarazo se debe adoptar como terapia de profilaxis en recién nacidos (para reducir las tasas de trasmisión madre-hijo); los ECA hechos contra placebo han demostrado una reducción en las tasas de trasmisión vertical, pero no hay estudios de seguimiento a largo plazo que evalúen la seguridad de dicho tratamiento en el niño; por tanto, no se conocen los efectos sobre el desarrollo neurológico o psicomotor a largo plazo; en este caso, conocer las preferencias de los pacientes (madre) y sus familiares (padre) acerca del tratamiento y de sus riesgos potenciales es un aspecto relevante para la decisión; sin embargo, no se identifican estudios cualitativos o cuantitativos sobre el tema. En este caso, puede ser particularmente necesario llevar a cabo investigación primaria para conocer las percepciones y las preferencias de los pacientes afectados o de la sociedad.

Algunos de los métodos que se pueden usar para tal tipo de investigación son: métodos cualitativos, como entrevistas personales y grupos focales de discusión, y métodos cuantitativos, como cuestionarios y encuestas[18,20,22].

Una vez se han completado los procesos de selección y evaluación de la información, de síntesis de la evidencia y de evaluación del cuerpo global de la evidencia, en cada uno de los dominios que contendrá el informe, los evaluadores pasan a generar las conclusiones sobre cada dominio en particular, así como un breve resumen de todas ellas, a manera de "puntos clave" para el tomador de decisiones.



El inicio de investigaciones en el contexto de la ET solo es pertinente cuando el conocimiento que se ha obtenido con la revisión de la literatura prueba ser inadecuado o insuficiente en los dominios más relevantes para la toma de decisiones.

Fase 4. Generación del informe preliminar que presenta los resultados de la ET

En esta fase, los resultados de la evaluación se organizan para dar estructura al informe. Al igual que con cualquier informe que presenta resultados de investigación, la presentación del informe debe estar enfocada en dar respuesta a los objetivos propuestos (resolver cada una de las pregunta planeadas), y, por tanto, tiene una relación directa con el protocolo. Los informes de ET deben ser balanceados y transparentes, y escritos, idealmente, en un lenguaje sencillo y claro, en el cual se eviten las repeticiones innecesarias. El texto debe estructurarse lógicamente y ser corto y preciso, pero, al mismo tiempo, deberá ser lo bastante detallado para que el lector pueda ver su fundamento y entender su metodología, así como reconocer los problemas metodológicos e interpretar el significado y la importancia de los resultados.

Es común hallar que el informe se estructura por capítulos (uno por cada dominio), en cada uno de los cuales se especifican las preguntas que se identificaron como relevantes acerca del uso de la tecnología para cada dominio en particular, los métodos seguidos para dar respuesta a esas preguntas y los resultados encontrados. Las limitaciones metodológicas deben ser claramente establecidas, así como el significado que estas pueden tener en la validez de los resultados.

Los resultados deben presentarse sin ambigüedades, e incluir tablas y figuras con los datos relevantes que complementen el texto. Con el fin de limitar la extensión del informe a solo los aspectos clave de mayor relevancia para el usuario del documento, los autores pueden informar los métodos detallados y los resultados de búsquedas, y evaluaciones o informes de las investigaciones realizadas para cada pregunta de investigación, en documentos de soporte anexos a la evaluación[5,19,20,31]. Si no se obtuvo información suficiente o adecuada para dar respuesta a alguno o a varios de los interrogantes planteados, es importante que eso quede descrito y fácilmente identificable dentro del informe; preferiblemente, en un aparte como "lo que no se conoce", pues dicha información puede guiar las agendas de investigación en los sistemas de salud.

Por último, todos los informes de dominio se toman juntos para generar el resumen global de los hallazgos, establecer el balance entre los beneficios y perjuicios que supone el uso de la tecnología, y definir las conclusiones, las cuales, a su vez, deben basarse en los resultados, sin sobredimensionarlos ni sobreinterpretarlos.

La **tabla 25.5** presenta los contenidos del informe que son comunes en los manuales de las asociaciones y las agencias de tecnología internacionales. Como ya se mencionó, los contenidos dependen del tipo de ET que se ha comisionado (completa, o breve); mientras la ET completa incluye todos dominios en la sección de informe de la evaluación, las ET breves pueden incluir solo los cuatro primeros de tal sección. Cabe aclarar también que la estructura del informe puede variar según el formato establecido por cada agencia, o por la entidad que comisiona la evaluación[31] (por ejemplo, el sistema de salud de Canadá tiene un formato prediseñado, el cual les entrega a las agencias que comisiona, para que en él incluyan la información que requieren sobre la tecnología).



El informe debe seguir una estructura lógica y ser corto y preciso, pero también lo bastante detallado para que el lector pueda ver su fundamento, entender su metodología e interpretar el significado de sus resultados.

Tabla 25.5 Contenido del informe de una ET

1. **Información general del proyecto**
   - 1.1. Título o nombre del proyecto de evaluación
   - 1.2. Número de identificación
   - 1.3. Tipo de evaluación (tipo de informe)
   - 1.4. Entidad que solicitó la evaluación de la tecnología (ET) en salud
   - 1.5. Autores y su afiliación
   - 1.6. Declaración de conflictos de interés
2. **Informe de evaluación**
   - 2.1. Descripción del problema de salud, y uso actual de la tecnología
   - 2.2. Alcance: Definido en términos de la pregunta que la ET responde en cuanto a:
     - Población: Población objeto de la tecnología
     - La intervención: Descripción y características técnicas de la tecnología
     - Comparación: Otras tecnologías disponibles para la indicación de interés
     - Resultados de salud: Lo que se espera modificar con el uso de la tecnología
   - 2.3. Seguridad
     - Preguntas
     - Descripción de la metodología seguida (fuentes de información utilizadas (búsqueda); proceso de selección y evaluación de la evidencia; síntesis y calidad del cuerpo de evidencia)
     - Resultados-documentos de soporte (tablas de resumen de hallazgos)
   - 2.4. Efectividad clínica
     - Preguntas
     - Descripción de la metodología (igual al anterior)
     - Resultados-documentos de soporte (igual al anterior)
   - 2.5. Costos y consideraciones económicas
     - Pregunta económica
     - Descripción del análisis económico completo
     - Resultados-documentos de soporte: tablas de información de efectividad y costos
   - 2.6. Consideraciones éticas
     - Preguntas
     - Descripción de la metodología
     - Resultados-documentos de soporte
   - 2.7. Aspectos sociales
     - Preguntas evaluadas
     - Descripción de la metodología
     - Resultados-documentos de soporte
   - 2.8. Aspectos organizacionales
     - Preguntas evaluadas
     - Descripción de la metodología
     - Resultados-documentos de soporte
   - 2.9. Resumen de los principales hallazgos (los relevantes para la decisión)
   - 2.10. Discusión-balance riesgo-beneficio
   - 2.11. Conclusiones para la decisión

Es necesario juntar los resultados de la evaluación en cada dominio, para establecer el balance entre los beneficios y los perjuicios que supone el uso de la tecnología, y definir las conclusiones del caso.

Fase 5. Revisión por pares-ajustes a la evaluación

Para asegurar su calidad, es deseable que el informe de la ET sea revisado al menos por un par externo antes de su publicación. El propósito de dicha revisión crítica por un igual es ayudar a los autores a escribir un manuscrito que sea fácil de leer y de alto nivel profesional. En consecuencia, durante esta fase se contrata a un experto que sea capaz de evaluar la metodología aplicada, la validez de los resultados, las conclusiones y las recomendaciones, así como la legibilidad y el alcance del manuscrito. Se espera que el evaluador haga una retroalimentación que ayude a mejorar el informe mostrando sus puntos fuertes y sus debilidades principales, así como sugerencias para los cambios que se deben hacer antes de su publicación. Finalmente, el diálogo entre los autores y el revisor permitirá hacer los ajustes necesarios mientras se mantiene el balance entre la claridad y la pertinencia de los contenidos con la extensión del informe[5].

Fase 6. Publicación de resultados

Una ET debe comunicarse activamente; sin embargo, la forma de diseminación y el tipo de publicación pueden variar dependiendo de lo que ha generado la evaluación y del interés de la entidad que la comisiona (generalmente, los ministerios de salud y los tomadores de decisiones en el sistema de salud). Comúnmente, los informes se difunden en formato de impresión para grupos predefinidos y se publican íntegramente en formatos electrónicos: por ejemplo, en el boletín electrónico de noticias o en la página web de la agencia que las hace o de la entidad que la comisiona. Cuando se quiere generar interés en los resultados de la ET y contribuir a que esta impacte en la mayor cantidad posible de decisiones, utilizar varios medios simultáneamente aumentará la probabilidad de que eso se logre. Las principales agencias de ET están indexadas en bases de datos de ET, la Biblioteca Cochrane y Trip Database[32].

Cuando se quiere generar interés en los resultados de la ET y contribuir a que esta impacte en la mayor cantidad posible de decisiones, se deben utilizar simultáneamente varios medios para su publicación.

CONCLUSIONES

La evaluación de tecnologías en salud hace un aporte único a los procesos de toma de decisiones en diferentes niveles del sistema de salud.

De acuerdo con el concepto tan amplio que tiene "tecnología" en el campo de la salud, los principios de la ET son aplicables para evaluar las posibles consecuencias no solo de las intervenciones médicas, sino también, de los programas de salud, las intervenciones organizacionales, e, incluso, de las reforma a los sistemas salud, ya que estas últimas pueden ser consideradas una intervención en el sistema de salud.

La ET ofrece una evaluación completa de los efectos potenciales sobre la salud y de las consecuencias para el sistema de salud, la economía y la sociedad de una tecnología que se va a introducir o a excluir del sistema de salud. Para cumplir correctamente dicha tarea, es necesario seguir un proceso multidisciplinario, en el cual la evidencia de diferentes aproximaciones de investigación sea considerada un insumo para la evaluación.

Referencias

1. Meskó B. 20 Potential technological advances in the future of medicine: Part I [internet]. 2013 [Citado 2014 ago. 14]. Disponible en: http://scienceroll.com/2013/08/01/20-potential-technological-advances-in-the-future-of-medicine-part-i/
2. Organización Mundial de la Salud. Informe sobre la salud en el mundo 2010. Financiación de los sistemas de salud: el camino hacia la cobertura universal [internet]. Ginebra: OMS;2010.
3. Kristensen FB, Lampe K, Chase DL, et al. Practical

tools and methods for health technology assessment in Europe: structures, methodologies, and tools developed by the European Network for Health Technology Assessment, EUnetHTA. Int J Technol Assess Health Care. 2009;(25Suppl 2):1-8.
4. E Unet HTA. HTA Core Model Handbook. version 1.4, 17 [internet]. 2012 [citado 2014 jul. 5]. Disponible en: http://meka.thl.fi/htacore/ViewHandbook.aspx
5. Kristensen FB, Sigmund H, editores. Health technology assessment handbook. 2nd ed. [internet] Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health; 2008. [citado 2014 jul. 5]. Disponible en: http://sundhedsstyrelsen.dk/publ/publ2008/mtv/metode/hta_handbook_net_final.pdf
6. Davies L, Drummond M, Papanikolaou P. Prioritizing investments in health technology assessment. Can we assess potential value for money? Int J Technol Assess Health Care. 2000;16:73-91.
7. Velasco-Garrido M, Kristensen F, Nielsen C, et al. Health technology assessment and health policy-making in Europe Current status, challenges and potential. The European Observatory on Health Systems and Policies; Observatory Studies Series No 14 [internet]. [citado 2014 ago. 15] Disponible en: http://www.euro.who.int/__data/assets/pdf_file/0003/90426/E91922.pdf
8. Institute for Quality and Efficiency in Health Care. General Methods/Versión 4.1 de 28 [internet]. Noviembre 2013. [citado 2014 ago. 15]. Disponible en: https://www.iqwig.de/download/General_Methods_4-0.pdf
9. National Institute for Health and Clinical Excellence. Guide to the single technology appraisal process. [internet]. Octubre 2009. [citado 2014 ago. 15]. Disponible en: https://www.iqwig.de/download/General_Methods_4-0.pdf
10. Canadian Coordinating Office for Health Technology Assessment (CCOHTA). Influencing decision- and policy-making using health technology assessment. Module 6. [internet]. [citado 2014 jul. 10]. Disponible en: http://www.albertahealthservices.ca/Researchers/if-res-hta-manual-module-6.pdf
11. National Information Center on Health Services Research and Health Care Technology (NICHSR) of the National Library of Medicine (NLM). HTA 101: Introduction to health technology assessment. [Internet] 2004. [citado 2014 ago. 15]. Disponible en: http://www.nlm.nih.gov/nichsr/hta101/hta101.pdf
12. Tecnología. Diccionario de la Lengua Española. 22ª ed. Madrid: Real Academia Española [internet]. 2001 [citado 2014 ago. 5]. Disponible en: http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=tecnología
13. HTAGlosary.net. [internet] Glosario ETS. Disponible en: http://htaglossary.net/inicio
14. Goodman C. Introduction to health care technology assessment. National Library of Medicine: National Information Center on Health Services Research & Health Care Technology (NICHSR); 1998.
15. Freeman RE, Reed DL. Stockholders and stakeholders: A new perspective in corporate governance. Calif Manage Rev. 1983;25:88-106.
16. Henshall C, Oortwijn W, Stevens A, et al. Priority setting for health technology assessment. Theoretical considerations and practical approaches. Priority setting Subgroup of the EUR-ASSESS Project. Int J Technol Assess Health Care. 1997;13:144-85.
17. Velasco-Garrido M, Busse R. Policy brief-Health Technology Assessment. European Observatory on Health Systems and Policies. [Internet]. 2009 [citado 2014 ago. 5]. Disponible en: http://www.euro.who.int/en/about-us/partners/observatory/policy-briefs-and-summaries/health-technology-assessment-an-introduction-to-objectives,-role-of-evidence,-and-structure-in-europe
18. Andersen S. Etiskeaspekterved MTV [Ethical Aspects of HTA]. En: Kristensen FB, Sigmund H, editors. MTV, sundhedstjemesteforskningogkliniskpraksis. [HTA, health science research and clinical practice.] Copenhagen: Danish Health Science Research Board and Danish Institute for Health Technology Assessment; 2000. p. 51-57.
19. HTA Core Model Handbook (version 1.5, 8 Apr 2014). [internet]. 2014 [citado 2014 ago.14]. Disponible en: https://meka.thl.fi/htacore/HTACoreModel_Handbook_2014-04-08.pdf
20. Busse R, Orvain J, Velasco M, et al. Best practice in undertaking and reporting health technology assessments. Int J Technol Assess Health Care. 2002;18:361-422.
21. Chassany O, Sagnier P, Marquis P, et al. The ERIQA group. Patient reported outcomes: the example of health-related quality of life – a European guidance document for the improved integration of health related quality of life assessment in the drug regulatory process. DrugInf J. 2002;36:209-38.
22. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology. 2008.
23. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [internet]. The Cochrane Collaboration, 2011. [citado 2014 ago. 14]. Disponible en: www.cochrane-handbook.org.

24. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
25. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
26. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
27. Hannes K. Critical appraisal of qualitative research [Internet]. En: Noyes J, Booth A, Hannes K, et al., et al. editors. Supplementary guidance for inclusion of qualitative research in cochrane systematic reviews of interventions. Chapter 4. Version 1. Cochrane Collaboration Qualitative Methods Group, 2011. [citado 2014 ago. 14]. Disponible en: http://cqrmg.cochrane.org/supplemental-handbook-guidance
28. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490-4.
29. GRADE Working Group. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. HolgerSchünemann Jan Brożek, Gordon Guyatt, Andrew Oxman, editors. [Internet] Octubre 2013. [citado 2014 ago. 14]. Disponible en: http://www.guidelinedevelopment.org/handbook/
30. Schünemann H, Oxman A, Higgins J, et al. Presenting results and 'Summary of findings' tables. En: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 Chapter 11. 2011 [citado 2014 ago. 14]. Disponible en: www.cochrane-handbook.org.
31. Velasco M, Perleth M, Drummond M, et al. Best practice in undertaking and reporting health technology assessments. Working group 4 report. Int J Technol Assess Health Care. 2002;18:361-422.
32. Sobrido M, González G, Cerdá M, et al. Estrategias para la difusión y diseminación de los productos de Evaluación de Tecnologías Sanitarias (ETS) [internet]. Madrid: Ministerio de Ciencia e Innovación; 2009. [citado 2014 ago. 14]. Disponible en: http://www.sergas.es/docs/Avalia-t/EstDifMemFinal.pdf.



SECCIÓN 6

Conducción de estudios

CAPÍTULO 26 Conducción y monitorización de estudios clínicos
Martha Beatriz Delgado Ramírez, María Ximena Rojas Reyes

CAPÍTULO 27 Adherencia
Álvaro J. Ruiz Morales, Claudia Marcela Granados Rugeles

CAPÍTULO

26

Conducción y monitorización de estudios clínicos

MARTHA BEATRIZ DELGADO RAMÍREZ • MARÍA XIMENA ROJAS REYES

INTRODUCCIÓN

La investigación clínica de buena calidad es esencial para el desarrollo de nuevas tecnologías en salud y para el avance del conocimiento. El diseño de estudios en las áreas clínica y comunitaria, así como su proceso de planeación y ejecución, es fundamental para garantizar la validez de los resultados y para dar soporte a las decisiones tomadas a partir de ellos. La conducción de investigación clínica varía según sean el propósito y la estructura del estudio, aspectos que dependen, en la mayoría de los casos, del investigador principal, y en algunos casos, del patrocinador del estudio.

Los investigadores clínicos pueden participar en el desarrollo de estudios de diferentes estructuras:

- Estudios multisitio[1]: En ellos, profesionales de la salud de diferentes instituciones son invitados a participar, y reciben información y entrenamiento sobre los métodos específicos de ejecución, control y monitorización preestablecidos, pero sin participar en el diseño, la planeación ni el análisis mismo del estudio (por ejemplo, en estudios propuestos y patrocinados por la industria farmacéutica, como experimentos clínicos de fase II y III para el desarrollo de medicamentos, o de fase IV para el control posmercadeo).
- Estudios clínicos multicéntricos: También son estudios multisitio, pero en ellos los coinvestigadores en cada sitio tienen participación en el diseño, la planeación, la conducción y el análisis de la información.
- Estudios clínicos de un solo sitio[1]: Corresponden a estudios de estructura menos compleja (por ser más pequeños), en los que el investigador planea la ejecución de acuerdo con sus objetivos y con el tiempo y los recursos disponibles.

En cualquiera de los casos anteriores, se requiere mantener altos estándares de calidad durante el desarrollo de todo el estudio, para obtener resultados válidos que den cumplimiento a los objetivos. Al conjunto de actividades encaminadas a lograr ese máximo nivel de calidad en cada aspecto se le conoce como ***conducción de un estudio***.

Si bien se han propuesto guías internacionales que regulan los aspectos éticos de la investigación, no hay la misma claridad para los aspectos de conducción. Como resultado de las dificultades que enfrentaban las autoridades regulatorias en Estados Unidos, Europa y Japón en estandarización de estudios sobre medicamentos, se planteó la necesidad de tener estándares en la conducción de esos estudios para garantizar que

se tuvieran las mismas reglas, y que los estudios conducidos en cualquiera de los países que se acogiera a esas normas pudieran ser aceptados en cualquiera de los otros.

Como respuesta a dicha necesidad, se creó la Conferencia Internacional de Armonización de Requerimientos Técnicos para el Registro de Medicamentos (ICH, por las siglas en inglés de International Conference for Harmonization for Technical Requirements for Registration on Pharmaceutical for Human Use), que se inició en 1997[2], entre representantes de autoridades regulatorias de los países participantes y de la industria farmacéutica.

Como resultado de la ICH, se logró la reunión (armonización) de las guías regulatorias preexistentes y se publicó la *Guía de buenas prácticas clínicas*, que proporciona un estándar unificado para el diseño, la conducción, el registro y el reporte de experimentos que involucren la participación de seres humanos. También describe los documentos esenciales de un estudio que permiten evaluar la conducción de una investigación clínica y de la calidad de los datos producidos.

Si bien este documento tiene una finalidad principalmente regulatoria en investigación de medicamentos para su registro, es suficientemente amplia para ser un modelo adecuado en la investigación que involucra a seres humanos y tiene aplicabilidad general, aunque no todos sus aspectos son adecuados para todos los estudios (en los descriptivos, poblacionales, epidemiológicos o simplemente observacionales analíticos, hay muchos aspectos que no son aplicables, y deja de contemplar otros que sí lo son).

En general, puede considerarse que las recomendaciones de la *Guía de buenas prácticas clínicas* son aplicables para la conducción de experimentos clínicos, aunque no se tenga la finalidad de producir información para registrar un medicamento. Es esencial que quien planee desarrollar estudios experimentales que involucren a seres humanos, o participar como investigador en ellos, conozcan los principios y sigan las recomendaciones de las guías de buena práctica clínica.

El propósito de este capítulo es presentar las pautas básicas para conducir y monitorizar con eficacia un proyecto de investigación; también: los principales aspectos que se deben considerar durante la planeación, la recolección de información, el análisis y el procesamiento de datos; las acciones y las herramientas que pueden utilizarse para asegurar la buena calidad de los datos, y el desarrollo exitoso de un estudio bien diseñado.

Si bien hay particularidades que se deben atender en cada diseño, hay aspectos aplicables a cualquier tipo de diseño, y cuyos objetivos son prevenir, detectar y corregir de manera oportuna inconvenientes en el desarrollo del estudio[1,3].

PLANEACIÓN DEL ESTUDIO

La conducción de un estudio clínico comienza una vez el investigador ha establecido la pregunta de investigación, trazado el plan del estudio —conocido comúnmente como el protocolo de investigación— y obtenido los recursos para su desarrollo.



Un protocolo claro, con lineamientos bien definidos, es el primer elemento que determinará el éxito de la investigación.

En el desarrollo del protocolo debe siempre incluirse un plan detallado para la ejecución del estudio, que tenga definidos con claridad los procesos de selección, evaluación y seguimiento de los sujetos del estudio, así como la asignación de actividades y las responsabilidades de cada una de las personas que participan en él. En la **tabla 26.1** se resumen algunos de los principales aspectos de la conducción del estudio que deben describirse con exactitud en el protocolo[4].

CONFORMACIÓN DEL GRUPO DE TRABAJO

El grupo de trabajo que se conforme depende de la estructura y del tamaño del estudio. En la **figura 26.1** se presenta una aproximación general a la conformación del grupo de trabajo y a sus vínculos durante la investigación.

Identificación de los investigadores

En algunas ocasiones, el protocolo de estudio es diseñado por una organización patrocinadora

Tabla 26.1 Principales aspectos de la conducción del estudio que deben describirse con exactitud en el protocolo

Seguimiento y manejo de pacientes	Administrativos
Duración de la inclusión de pacientes y del seguimiento.	Responsabilidades administrativas del investigador principal y del grupo coordinador.
Ubicación del estudio (lugar, centro coordinador, centros participantes).	**Aprobaciones institucionales.**
Procedimientos para el envío de material del estudio desde y hacia el centro coordinador.	**Estrategias para mantener la confidencialidad de los** datos.
Obtención, manejo y envío de especímenes de laboratorio, etc.	Especificaciones sobre recolección y procesamiento de datos.
Definición del momento en el que se incluirá a los pacientes.	Plan de control y de monitorización.
Informe y clasificación de eventos adversos.	Manejo de enmiendas al protocolo.
Controles de seguimiento como tiempo, falta de cumplimiento.	Autoría y publicación de los datos.
Pérdidas del seguimiento de pacientes.	
Retiro de pacientes y descontinuación temprana del paciente en el estudio.	
Manejo de emergencias y de eventos adversos serios.	

Fuente: Elaboración propia de las autoras.

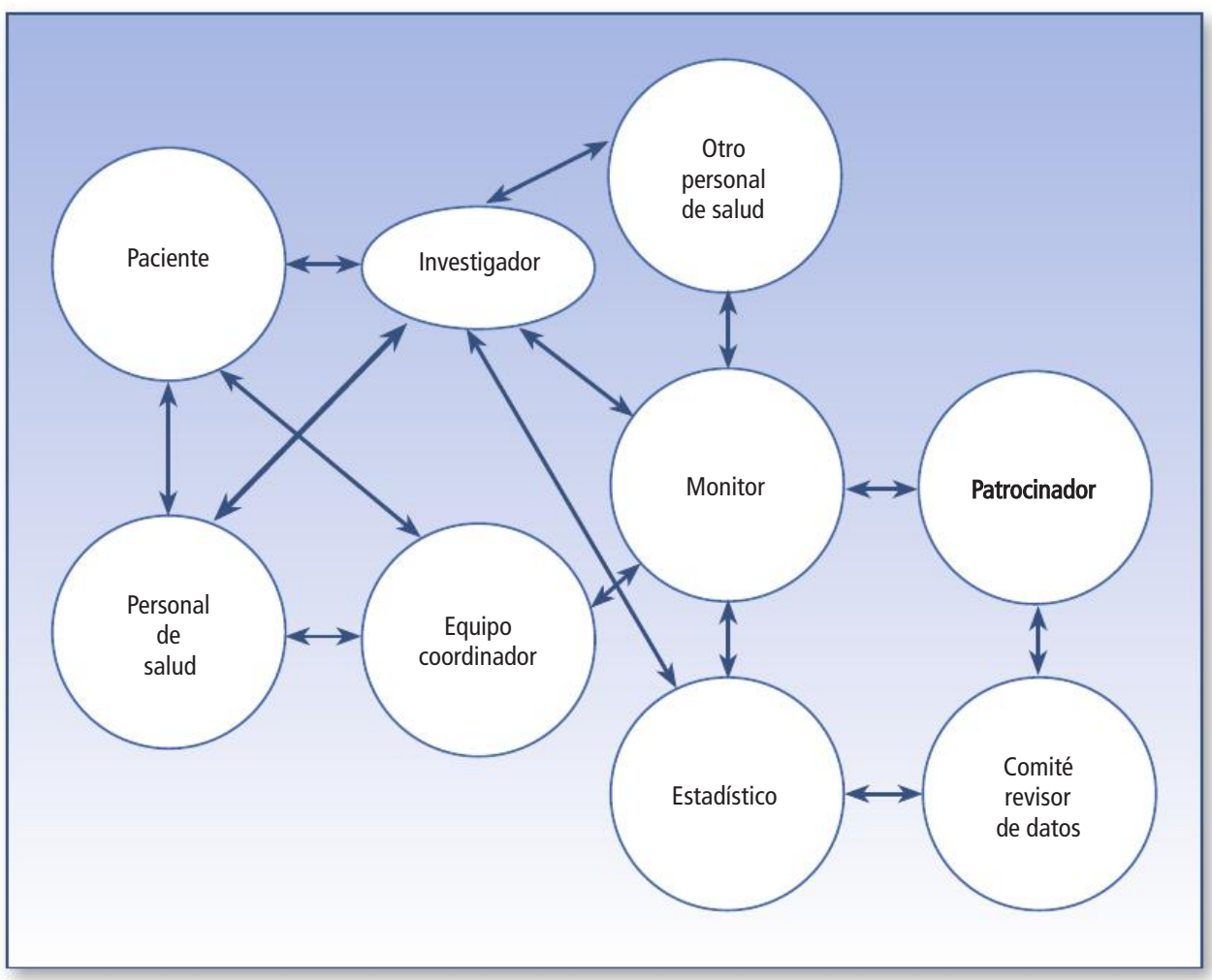


Figura 26.1 Equipo de trabajo en una investigación clínica

que no cuenta por sí misma con el personal que lo llevará a cabo.

En tales circunstancias, el patrocinador debe seleccionar a los investigadores que participarán en el estudio

Según aspectos como la amplia experiencia clínica e investigativa, que les permitan proveer soporte médico durante toda la conducción del estudio, capacidad técnica para ejecutar un estudio, acceso a los sujetos que el estudio requiere y capacidad para conseguir la aprobación del estudio ante comités de ética, entre otros[5].

Funciones del investigador

En todas las fases del estudio, el investigador tiene responsabilidades, las cuales se citan a continuación:

- Antes de iniciarse el estudio, debe demostrar su propia idoneidad como profesional para el manejo de los sujetos del estudio; debe, además, proveer información sobre su hoja de vida, demostrar que tiene a su alcance los recursos técnicos apropiados para el manejo de los sujetos y que se hará responsable de la asistencia médica de los pacientes del estudio. Adicionalmente, debe mantener el contacto con el comité de ética y conseguir de este la autorización para llevar a cabo el estudio[2].
- Durante la ejecución y el reclutamiento de los sujetos **(tabla 26.2), debe comprometerse** a mantener total adherencia al protocolo y a dar un adecuado manejo a los productos que se estén utilizando en la investigación. En caso de hallarse conduciendo un experimento clínico, es responsable del consentimiento informado que debe firmar cada paciente en el estudio, bien sea en persona o por intermedio de alguien delegado por él; debe conocer con claridad el procedimiento de aleatorización y el de apertura del enmascaramiento o el cegamiento, así como asegurar un adecuado seguimiento a todos los sujetos en el estudio; en particular, en los aspectos relacionados con la evaluación de la seguridad[2,6].

Conformación de comités externos

Para la ejecución de estudios experimentales que evalúan medicamentos o nueva tecnología, es necesario, en ocasiones, conformar un comité independiente de los investigadores para evaluar los datos del estudio, y el cual tiene como función vigilar la información acumulada para detectar oportunamente el beneficio de la intervención o sus efectos adversos durante el progreso del estudio[6,7].

La conformación de estos comités externos de evaluación es vital cuando se planean análisis interinos, para evaluar la seguridad de la intervención (efectos adversos) o cuando se planea un seguimiento a muy largo plazo y el desenlace medido es la mortalidad o una morbilidad significativa[6,7].

Los miembros de dichos comités deben ser personas expertas en el tema de estudio, e independientes del patrocinador y de los investigadores.

Tabla 26.2 Funciones del investigador durante el reclutamiento de sujetos

1. Proveer cuidado médico a los pacientes en estudio.
   - Demostrar cualificación para toma de decisiones médicas.
   - Garantizar cuidado médico adecuado ante eventos adversos.
   - Mantener el listado de personas delegadas para realizar procedimientos del estudio.
2. Mantener la confidencialidad.
   - Proteger la identidad del sujeto experimental, que debe ser confidencial durante el estudio y después de este.
   - Vigilar que el nombre del sujeto no aparezca en el formato de reporte de caso (FRC), así como tampoco datos que permitan su identificación.
3. Adherirse al protocolo y mantener los documentos fuente.
   - Reclutar a los pacientes adecuados.
   - Seguir los procedimientos de aleatorización.
   - Hacer las evaluaciones de seguimiento.
   - Justificar cualquier desviación del protocolo.
4. Manejar el medicamento de estudio.
   - Conservar y usar el medicamento de acuerdo con el protocolo.
   - Asegurar su almacenamiento adecuado.
   - Mantener actualizado el inventario de la medicación.
   - Asegurar el uso correcto del medicamento por parte de los sujetos experimentales.
5. Evaluar la seguridad de los sujetos del estudio.

Fuente: Elaboración propia de las autoras.

CONSIDERACIONES ANTES DE DAR COMIENZO AL ESTUDIO

Evaluación de factibilidad del protocolo

Uno de los primeros aspectos que debe evaluar el investigador es si el estudio será factible con el protocolo que se planeó. Dicha evaluación, en la planeación, no puede ser pasada por alto, y debe evitarse que un investigador entusiasta idealice la factibilidad de su propio estudio, lo cual le traerá dificultades en la ejecución y el análisis, y, más aún, puede hacer que el proyecto fracase. Los aspectos más relevantes que se deben evaluar se enumeran en la **tabla 26.3**.

- Identificación de la población de estudio:

El investigador principal debe identificar a la población de estudio de acuerdo con los criterios de admisión del protocolo y evaluar de antemano la probabilidad del reclutamiento de sujetos.

Además, debe tener en cuenta el área y la población de la cual extraerá la población de estudio. Debe determinar la frecuencia con la que se han registrado en los últimos meses los casos de interés, la disposición de los sujetos para participar, la posibilidad de aceptación de las pruebas requeridas por el protocolo, etc. La consideración de cada uno de estos aspectos permitirá identificar las mejores fuentes de reclutamiento, los factores poblacionales, del medio y del protocolo mismo que pueden interferir en la consecución de la muestra, de manera que se puedan adoptar desde el comienzo las medidas preventivas y correctivas necesarias[8,9].

Tabla 26.3 Aspectos para evaluar la factibilidad del protocolo

Fuentes de sujetos/Identificación de la población de estudio.
Interés de los investigadores en cada centro, y la de los colaboradores.
Infraestructura requerida para el desarrollo del estudio en los centros participantes.
Tiempo de duración del estudio: • Tiempo requerido para la inclusión de pacientes. • Periodo de seguimiento de cada individuo participante. • Tiempo de procesamiento y análisis de la información. • Tiempo requerido para manuscritos e informes finales.

Fuente: Elaboración propia de las autoras.

- Conocimiento e interés de los investigadores y colaboradores: **Es deseable que los coinvestigadores** y los colaboradores que se requieran para el estudio sean líderes académicos en el tema del estudio, que tengan interés, e, idealmente, que hayan tenido experiencia en investigación.

Cuando deba involucrarse a más de una institución, es indispensable asegurarse de lograr la máxima cooperación, la cual, depende, fundamentalmente, de la motivación, el conocimiento y el entrenamiento que hayan recibido. Debe considerarse si el tema es atractivo, si los procedimientos de evaluación y seguimiento propuestos son factibles y aceptados y si conocen el manejo seguro del tratamiento o la terapia experimental.

La falta de aceptación de otros colegas y colaboradores puede introducir sesgos por fallas en el cumplimiento de los procedimientos del estudio en alguna de las instituciones, lo cual puede invalidar los resultados[8].

Para lograr la máxima cooperación, debe ofrecerse información en reuniones previas, dirigidas a todos los que tendrán contacto con los sujetos del estudio. Tal estrategia le permite al investigador evaluar el conocimiento y la aceptación que se tenga del protocolo por parte del equipo; también, conocer los procesos existentes y las pautas de manejo establecidas en cada centro para el tratamiento de los casos de interés para el estudio, e identificar los aspectos que requieren ajuste o modificación para lograr los objetivos de la investigación.

- Infraestructura requerida para el desarrollo del estudio: **Es importante visitar las instituciones,** conocer la infraestructura con la que cuentan y evaluar la disponibilidad de los requerimientos tecnológicos del protocolo, como un paso de evaluación de factibilidad.
- Tiempo de duración del estudio: **Cuánto dure,** aproximadamente, la fase de recolección

depende del tamaño de muestra, de la tasa de reclutamiento en cada institución y del tiempo de seguimiento. Para dicho cálculo previo, es útil conocer las estadísticas de las instituciones participantes[9]. Tales consideraciones permiten determinar el número de personas necesarias para recolectar información, así como los recursos financieros requeridos.

Proceso de aprobación del estudio. Entes reguladores y comité de ética

Contar con la aprobación del comité de investigación y ética de cada institución es un paso previo indispensable para dar inicio al estudio.

Hay unas recomendaciones estandarizadas para la evaluación por parte de los comités de ética e investigación, las cuales los investigadores deben conocer (Declaración de Helsinki, Guías de Buenas Prácticas Clínicas, y Guías del Consejo de Organizaciones Internacionales de las Ciencias Médicas —CIOMS, por las siglas en inglés de Council for International Organizations of Medical Sciences—) para investigación biomédica). Adicionalmente, cada comité de ética e investigación puede tener algunos criterios adicionales, los cuales el investigador debe conocer para asegurarse de que han sido contemplados en el protocolo[10,11].

En Colombia, mediante la Resolución 002378 de 2008, se adoptan las *buenas prácticas clínicas* para las instituciones que conducen investigación con medicamentos en seres humanos[12].

Aquí se presentan, en términos generales, algunas consideraciones que se toman en cuenta en la evaluación de propuestas de estudios experimentales **(tabla 26.4)**, que se amplían en el **capítulo 15** "El experimento clínico", de este mismo libro. Es importante que el investigador conozca de antemano los requisitos y los documentos requeridos por cada uno de los comités de ética e investigación para presentar propuestas.

Planeación de aspectos logísticos y administrativos

El investigador principal y el coordinador del estudio deben planear cuidadosamente los aspectos logísticos del proyecto.

Estos varían según la estructura y el tamaño del estudio, el número de centros participantes, la vinculación de asistentes de investigación y los recolectores de información, la ubicación del lugar donde se harán y se analizarán los exámenes paraclínicos, etc. La **tabla 26.5** muestra algunos aspectos generales que se deben establecer antes del desarrollo del protocolo[13,14].

PREPARACIÓN PARA RECOLECTAR INFORMACIÓN

Diseño de los formatos de recolección

Para elaborar los formatos de recolección, debe haberse definido cuáles son las variables según las cuales se recolectará información, bien sea para evaluar los criterios de inclusión, bien sea para el control durante las visitas, o bien, para el análisis final.

Debe tenerse en cuenta la unidad de medida de cada variable, la relación entre ellas, en cuáles puede requerirse información adicional y el tiempo que será necesario para su medición. En la **tabla 26.6** se enumeran los componentes básicos de un formato para recolectar información.

Hay descritos en la literatura múltiples diseños para desarrollar formatos de recolección, y su selección varía según el estudio y la forma como se recogerán los datos (entrevista con pregunta abierta, cuestionarios estructurados, formatos para autodiligenciamiento o para recolectar información a partir de otros registros como historias clínicas, reportes de exámenes paraclínicos, etc.).

Los instrumentos usados más a menudo son los cuestionarios, aunque hay otros, como la observación, los archivos, las escalas de medición,

Tabla 26.4 Consideraciones éticas en la evaluación de propuestas de estudios experimentales

¿Cuál es la necesidad real de llevar a cabo la investigación?
¿Es apropiado el uso de placebo en el grupo control?
En caso de tener un comparador activo, ¿se trata de la mejor opción disponible?
¿Se ha seleccionado la terapia adecuada en el control activo?
¿Se han considerado y manejado los posibles sesgos?
¿Se ha hecho una adecuada consideración del tamaño de muestra requerido?
¿El protocolo tiene un poder estadístico suficiente?
¿La probabilidad de recibir el tratamiento experimental y de control es igual para todos los pacientes incluidos en el estudio?
¿Cómo es la seguridad para los pacientes en estudio?

Fuente: Elaboración propia de las autoras.

Tabla 26.5 Aspectos logísticos que requieren planeación

1. Sistema de pago al personal que colabora en el estudio.
2. Distribución, control y almacenamiento de los medicamentos y de los demás materiales del estudio en cada centro.
3. Toma, manejo, conservación y transporte (si se requiere) de especímenes para análisis.
4. Formas de envío o de entrega de formatos de registro al centro de manejo de datos.
5. Periodicidad con la que se requiere reinstrucción en los procedimientos del estudio.
6. Pagos o manejo de incentivos o de compensaciones que se darán a los participantes por su asistencia a controles y a las demás citas requeridas por el estudio.
7. Periodicidad y lugar de reuniones del grupo investigador y de los comités de revisión.

Fuente: Elaboración propia de las autoras.

los formatos para recolectar datos de laboratorio, las técnicas proyectivas, los grupos focales, las entrevistas en profundidad, etc. Independientemente del tipo de instrumento escogido para la investigación en particular, debe garantizarse que sea autoexplicativo y comprensible, lo cual permite mayor confianza en la validez de la medición y puede simplificar el trabajo de los recolectores de los datos[15-18].



Cualquiera que sea el instrumento de recolección usado, debe haber sido coordinado con el sistema de manejo de los datos[19], con el diseño de la base de datos y con el plan de análisis.

Por ello, se requiere un estrecho trabajo en equipo entre el investigador principal, el coordinador y el bioestadístico, con el fin de identificar las variables que serán medidas directamente y las que serán generadas a partir de estas[1].



En los estudios multicéntricos, es aconsejable el uso de formatos estructurados; en lo posible, con preguntas cerradas y respuestas codificadas.

Esto permite mantener unificados los criterios de medición de las variables y facilita el procesamiento y el análisis de datos. Cada hoja o formato debe tener un espacio para la información de identificación del paciente en el estudio, lo que ayuda a disminuir el riesgo de pérdida de información[1,20].

Debe haber instrucciones sobre cómo hacer el registro de la información en cada hoja de recolección, lo que facilita su diligenciamiento, permite unificar el registro entre los recolectores y evita inconvenientes durante el proceso de digitación.

Es aconsejable que tanto los investigadores como el monitor y el bioestadístico hagan varias revisiones de los instrumentos de recolección antes de empezar a recolectar datos, así como

Tabla 26.6 Componentes básicos de un instrumento de recolección

1. Información general. • Número de identificación del individuo dentro del estudio. • Características sociodemográficas (sexo, edad, raza, nivel educativo, estado civil, etc.). • Lugar o centro donde se recluta al paciente. • Fecha de su inclusión en el estudio. • Nombre del encuestador.
2. Información sobre las variables que se van a evaluar.
3. Instrucciones sobre la medición de cada variable, sobre el registro de información y sobre frases indicadoras de salto y orden, etc.

Fuente: Elaboración propia de las autoras.

hacer una prueba piloto del instrumento, para corregir los defectos, mejorar los formatos y tener una idea del proceso y de su duración.

Preparación del sistema de aleatorización al tratamiento

Como se mencionó en capítulos anteriores, la asignación aleatoria a las intervenciones evita sesgos y asegura la comparabilidad de los grupos, por lo que la planeación cuidadosa del sistema de asignación es de vital importancia.

La preparación de la aleatorización debe asegurar que el clínico no conocerá la asignación antes del inicio del tratamiento[20,21], para lo cual hay varias estrategias, como el uso de sobres opacos sellados, la asignación del tratamiento desde un sitio remoto (el centro coordinador) y el uso de tablas de números aleatorios.

El investigador debe asegurarse de que los participantes respeten la asignación que les dé el azar. Es recomendable conservar una copia de la aleatorización en el centro coordinador, en caso de que se requiera verificarla.

Preparación de la terapia o el tratamiento experimental-Enmascaramiento

Una vez se ha establecido cuáles serán los tratamientos experimental y de control, el investigador debe planear la mejor forma de enmascarar las terapias para mantener el ciego en los estudios que así lo requieran. En algunos casos, es más difícil, por las características de la intervención: por ejemplo, un procedimiento quirúrgico o una terapia pueden presentar dificultades en el enmascaramiento.

En la conducción del estudio, es vital conocer los métodos para enmascaramiento, haberlos probado previamente y seguirlos, para evitar que se conozca la terapia[22].

En caso de experimentos con medicamentos, deben conocerse la dosis requerida y la frecuencia de administración, así como los detalles relacionados con la preparación del medicamento y la estabilidad de las mezclas, antes de determinar la forma de enmascararlos.

Si la terapia de control es un placebo, deberá verificarse que este es idéntico al medicamento activo experimental en aspectos como la apariencia general, la superficie y el color interior, el tamaño, el sabor, el olor, el peso, los grabados o las impresiones sobre la superficie, la viscosidad y las texturas interna y externa.

Solucionados dichos aspectos, se debe planear la forma como serán empacados y distribuidos los tratamientos, bien sea directamente al paciente, o bien, a los centros investigadores, para su aplicación por parte del personal de salud.

Entre otras condiciones que deben prepararse, están el tipo de recipiente o de empaque, cómo utilizar los códigos de aleatorización, la cantidad de medicamento en cada empaque (por dosis, por unidad…), las preparaciones extras en caso de extravío y la posibilidad de aumento o reducción de dosis durante el estudio, entre otras[22].

Estandarización de mediciones y estrategias de recolección

Talleres de entrenamiento de participantes en el estudio

Es claro que el éxito del estudio dependerá de la calidad de la información recogida para su análisis, por lo cual la selección y el entrenamiento de las personas que harán esa tarea son de vital importancia.

Las habilidades de los encuestadores o los recolectores de información pueden tener un impacto sustancial sobre la calidad de las respuestas. Bourque y Clark[15] describen las ventajas de contar con encuestadores expertos y los riesgos de quienes no cuentan con experiencia. Es importante que se determinen durante el proceso de selección de los encuestadores o los asistentes de recolección sus motivaciones, sus expectativas laborales y si cumplen el nivel educativo requerido para la tarea que van a desarrollar, al igual que otros aspectos

importantes, como su presentación personal y su experiencia en actividades similares.

Con el entrenamiento de los encuestadores y los recolectores, se pretende unificar los criterios de recolección entre los diferentes asistentes y los centros de investigación, hacerles explícita la importancia de su papel dentro del estudio y motivarlos para el uso de las técnicas correctas de recolección, como no guiar las respuestas, utilizar correctamente los manuales de procedimientos y el del instrumento de recolección, y mantenerse en permanente contacto con el coordinador local o general del estudio[23].

El coordinador del estudio y el investigador deben planear por anticipado los temas que se tratarán durante el entrenamiento, así como identificar a las personas encargadas de cada sesión y la metodología que se va a seguir.

La técnica más recomendada es hacer talleres de entrenamiento con una parte explicativa sobre todo el estudio, seguida por ejercicios en los cuales se practique el diligenciamiento de formas de recolección.

Durante la primera parte explicativa se recomienda presentar los objetivos del estudio, su justificación, las fuentes de financiación, la metodología empleada (incluyendo el tipo de muestreo empleado), una discusión de los aspectos éticos del estudio, el anonimato de la información, las fuentes de sesgos, las características de los instrumentos de recolección, la forma como se deben diligenciar los instrumentos de recolección, una presentación de las ventajas y las desventajas de seguir las instrucciones, y la organización del trabajo, ya sea en campo o por teléfono, o en el hospital, etc.

A la parte práctica se recomienda dedicarle el mayor tiempo posible practicando las formas de diligenciar los instrumentos y usando los manuales del instrumento correspondiente, hasta confirmar el pleno entendimiento por parte de todos los asistentes[8].

Al finalizar el taller, se recomienda entregar tareas específicas para llevar a cabo durante los días siguientes al entrenamiento, con el fin de identificar nuevas inquietudes y vacíos no resueltos, para resolverlos en una segunda reunión antes de dar inicio a la fase de recolección del estudio. El coordinador debe estar dispuesto a hacer cuantas reuniones aclaratorias requieran los asistentes para resolver dudas y afianzar las funciones esperadas[24].

Estudio piloto

El estudio piloto es necesario para identificar problemas durante la recolección, así como para el reconocimiento del área de estudio, para estimar la tasa de reclutamiento real, para determinar el tiempo requerido para el diligenciamiento de los instrumentos, e, igualmente, para evaluar el contenido de forma y fondo de los instrumentos. De manera simultánea, se tiene la oportunidad de evaluar el desempeño de los asistentes de recolección en sus funciones y verificar el sistema de manejo de datos[19,25]. En algunos casos, se lo usa para determinar la frecuencia de aparición de un desenlace cuando no hay información previa o información que permita el cálculo apropiado del tamaño de muestra en un estudio posterior.

Durante esta etapa, el coordinador del estudio juega un papel importante a la hora de asignar las funciones de los asistentes y en el seguimiento del proceso de recolección y de identificación de las posibles dificultades.

Con los resultados obtenidos en dicho estudio, se espera hacer las correcciones necesarias y los ajustes a los cálculos de tiempo y de costo, hacer un reentrenamiento a los recolectores en los puntos donde se identificaron fallas, mejorar el sistema de muestreo o recalcular el tamaño de muestra, etc.

Elaboración de los manuales de procedimientos

El coordinador del estudio, de común acuerdo con el investigador, debe darse a la tarea de elaborar las pautas requeridas para dar cumplimiento al protocolo de estudio, pautas que se deben agrupar en un manual de funciones y procedimientos, dirigido a los monitores o los coordinadores de cada centro, así como a las personas encargadas de la recolección y el manejo de esta[26].

En esencia, los mencionados manuales presentan con claridad las acciones que se deben desarrollar durante el estudio, y no deben ser muy extensos ni complicados, para que se puedan consultar fácilmente y no terminen siendo archivados en los centros de investigación por su complejidad y poca funcionalidad.

Un ejemplo de la información que pueden contener estos manuales se presenta en la **tabla 26.7.**

Registros de control de los procedimientos durante el proceso de recolección

- Registro de tamizado y reclutamiento de pacientes: Este registro permite conocer el número de pacientes evaluados como elegibles para ingresar al estudio, y cuáles son las razones por las cuales algún paciente no fue reclutado. Dicha información será de utilidad para identificar las razones de baja aceptación o de bajas tasas de reclutamiento.
- Registro de control de medicamentos: Cuando el estudio involucra la entrega de medicamentos al paciente para su autoadministración, dicho registro es indispensable para controlar la distribución de los medicamentos, y puede servir también como herramienta para medir el grado de adherencia del paciente al tratamiento. En él se registra: el total de medicamento entregado (por ejemplo, el número de tabletas), la fecha de entrega, la cantidad de medicamentos devueltos por el paciente, la fecha programada para una nueva entrega y el nombre del asistente de investigación que entrega y recibe los medicamentos.

Tabla 26.7 Información que debe contener el manual del estudio

1. Resumen del protocolo, con énfasis en:
• Objetivos del estudio.
• Justificación.
• Lugares en los que se desarrollará la investigación.
• Población de estudio.
• Tipo de muestreo.
2. Procedimientos para:
• Reclutamiento de los individuos.
• Selección de la muestra.
• Recolección de la información.
• Manejo de las muestras y de los especímenes.
3. Sistema de comunicación, manejo y control de información entre los centros participantes y el centro coordinador.
4. Funciones y responsabilidades de cada miembro del grupo investigador y de los colaboradores.
5. Directorio del estudio.

Fuente: Elaboración propia de las autoras.

Planeación del proceso y análisis de datos

Las actividades en esta fase de la conducción dependen del sistema utilizado para recolectar la información, sea este centralizado o no[19,27].

Durante la fase de recolección de información, es básico que el coordinador designe a una persona encargada de revisar todos los instrumentos de recolección una vez diligenciados, antes de ingresar la información al programa de captura o a la base de datos[19,27].

Cuando se tiene un sistema de manejo *centralizado,* durante esta fase se busca información incompleta o inconsistente, así como errores al diligenciar los instrumentos de recolección; además, se revisa si la información falta porque el original es ilegible, o porque hay espacios en blanco, y, en este caso, se marca como faltante; se evalúa la concordancia entre los datos de identificación del centro, sus formatos y el número de identificación de los pacientes, la concordancia de los datos con la clínica del paciente, la presencia de datos anormales o de valores insólitos en los resultados de exámenes paraclínicos, etc.[8].

Cuando el sistema de manejo de los datos es *no centralizado*, los coordinadores generales y locales, junto con los monitores, juegan un importante papel de revisión local con el bioestadístico encargado del estudio. En tales casos, la información puede digitarse en cada centro, y el centro coordinador tiene la responsabilidad de revisar las posibles inconsistencias[19,23].

Recientemente, se han incorporado en la práctica registros en línea de reporte de caso,

que permiten la captura directa de los datos para el análisis, lo cual, a su vez, minimiza el recurso requerido en este proceso de revisar la información y los errores potenciales en la digitación de los datos[27].

INICIO DEL ESTUDIO-PROCESO DE RECOLECCIÓN DE INFORMACIÓN

Durante este periodo de la investigación, el papel del coordinador del estudio se intensifica, y es aquí cuando se requieren apoyo y esfuerzo mutuos entre el investigador y el coordinador, para mantener los estándares deseados.

Algunas de las funciones del coordinador durante este periodo ya se han mencionado, e incluyen:

a) **Coordinar la recolección de la información** en los diferentes centros asignando funciones y tareas de recolección para cada asistente de investigación.
b) **Mantener contacto permanente con los asistentes** de recolección para resolver inquietudes, así como para dar apoyo y motivar.
c) **Visitar periódicamente cada centro de investigación**, con el fin de evaluar la adherencia al protocolo de los investigadores y de los asistentes de recolección, verificar las técnicas de muestreo y los demás procedimientos, y para revisar los formatos de seguimiento de recolección para detectar tempranamente errores e identificar el grado de veracidad de los datos. La frecuencia y la periodicidad de las visitas dependen de los problemas identificados en cada centro, así como también, de la fase en que la se encuentre el estudio.



Durante la fase inicial de recolección, se requerirá más apoyo por parte del coordinador y ayuda en solución de inconvenientes, y durante la fase intermedia se requerirá mantener la motivación hasta el final.

Se puede optar por diferentes técnicas de visita, pero es esencial evitar la imagen de policía o de supervisor, y debe lograrse empatía con el asistente o los asistentes, para una mayor cooperación con el proceso.

En lo posible, cada visita debe incluir una reunión con todo el grupo participante en ese centro de investigación, y es importante hacer énfasis en los objetivos de la visita, así como informar sobre el estado del estudio, lo que permitirá mantener el interés de los colaboradores.

d) **Establecer nuevas formas para el envío de** los instrumentos de recolección al centro coordinador (sistema de manejo centralizado)[19] o de la información, en caso de que el sistema de manejo del proyecto no sea centralizado.

Inclusión de pacientes

Fuentes de individuos

Existen diversas fuentes para reclutar a los pacientes, de acuerdo con los objetivos propuestos y los criterios de admisión[16]:

a) **Contactos con colegas que trabajen en el área** clínica sobre la cual se va a investigar, ya sea en consulta privada, en consulta externa o en centros de atención especializada.
b) **Contactos con centros de diagnóstico, tales** como laboratorios clínicos o de imágenes diagnósticas, o laboratorios de fisiología, de neurología, de cardiología, de gastroenterología, etc.
c) **Contactos con la industria privada.**
d) **Contactos con sectores específicos de la** población, como empleados del gobierno, militares, estudiantes universitarios, bachilleres o de primaria, reclusos, jugadores de determinado deporte, etc.
e) **Pacientes terminales.**
f) **Sujetos voluntarios**[28].

Para lograr un buen reclutamiento, se requieren creatividad, constancia y paciencia. Debe tenerse en cuenta, además, que no todos los posibles centros colaboradores ni todas las áreas de tamizado estarán de acuerdo con el protocolo, y, por tanto, es importante que el grupo investigador les venda la idea a los estamentos y los colegas de quienes requiere colaboración.



La promoción del estudio se puede lograr más fácilmente mostrando una investigación bien planeada, un diseño adecuado y una metodología clara, así como objetivos bien definidos y viables, y, en general, con una organización en la que se perciban solidez y profesionalismo.

Factores que influyen en el reclutamiento

Algunos de los aspectos que se deben tomar en consideración durante el reclutamiento son los relacionados con el tiempo y la disponibilidad de los participantes y las distancias entre su hogar o su sitio de trabajo y el lugar del estudio; también, los relacionados con el interés o las razones para participar (conciencia y relevancia de la enfermedad), su composición socioeconómica, el número total de individuos contactados, la necesidad de exámenes desagradables, incómodos, dispendiosos o que consumen mucho tiempo, y, sin duda, el tipo de compensación o de ganancias secundarias por participar activamente en el estudio (dinero, pago de transporte, estímulos, etc.)[28].

Seguimiento de la tasa de reclutamiento



Durante la conducción del estudio, es esencial llevar un registro sobre el proceso de reclutamiento en cada uno de los sitios de investigación, que se inicia identificando a la población potencial disponible en el área tamizada, a los sujetos disponibles para la entrevista inicial, a quienes no ingresaron al estudio por no cumplir criterios de admisión, a quienes aceptaron ingresar al estudio, a quienes firmaron el consentimiento informado, a quienes ingresaron a la fase de enriquecimiento y, por último, a quienes reciben tratamiento.

El mencionado balance es necesario en esta fase de inicio, así como también durante todas las visitas de seguimiento contempladas por el protocolo, y la visita final en casos pertinentes. Dichos balances de reclutamiento y la identificación de las razones por las cuales un paciente no progresa dentro de ellas son de gran utilidad durante el análisis de los datos y la conducción del estudio, pues permiten detectar oportunamente cualquier disminución en la tasa de reclutamiento esperada, con el fin de tomar a tiempo las medidas correctivas del caso[9,28].

Técnicas para aumentar el reclutamiento de pacientes

Durante el inicio de la investigación, es importante divulgarla en reuniones científicas de interés, tales como congresos, conferencias, simposios, etc. Si se considera necesario, se pueden editar volantes informativos para repartirlos dentro de la población potencial, poner avisos informativos en sitios estratégicos y hacer propaganda en prensa, radio y televisión local, según el caso.

Durante el proceso de reclutamiento se debe mantener contacto directo y periódico con los diferentes colaboradores, para conocer el desarrollo del estudio en su área de recolección e informarlos sobre el proceso del estudio en otros centros, así como para reconocer e incentivar su participación.



Ante una baja tasa de reclutamiento, se deben tomar con rapidez medidas correctivas.

El primer paso será identificar el origen del problema: si se debe a los sujetos, a los recolectores, a las exigencias del protocolo, a otros factores, o, tal vez, a una combinación de varios de ellos[8,9].

Como medidas correctivas, se pueden intensificar las acciones ya enunciadas para divulgar la investigación, considerar la posibilidad de aumentar el número de centros participantes en el estudio, visitar los centros de recolección para resolver dudas y definir nuevas acciones donde se canalicen esfuerzos que mejoren el reclutamiento, determinar cuáles medidas de reclutamiento están siendo más exitosas que otras y optar por ellas, aumentar la remuneración u otros beneficios secundarios a los recolectores o las compensaciones o los estímulos a los sujetos (siempre ciñéndose a los más estrictos principios éticos), contratar a un mayor número de recolectores, dejar abierta la posibilidad de que el sujeto siga recibiendo tratamiento (si es el caso, después de terminado el estudio), modificar los criterios de admisión —si es pertinente— o aumentar la duración del estudio[28].

Estos tres últimos recursos requerirán aprobación por parte del comité veedor, del comité de vigilancia y seguridad, del comité de investigación y ética, y algunas veces, de las fuentes financiadoras del estudio, así como una evaluación con el bioestadístico a cargo del estudio, para determinar los posibles efectos en la investigación.

El aumento en el tiempo de reclutamiento puede repercutir en efectos negativos, que deben ser considerados y manejados muy de cerca por

el coordinador del estudio, con el fin de evitar un fuerte impacto a lo largo del curso de la investigación.

La pérdida de motivación por parte de los colaboradores ante una baja tasa de reclutamiento puede llevar a la violación de los criterios de inclusión, implicar que se modifiquen los criterios diagnósticos aplicados a los sujetos a través del tiempo y generar la inclusión de diferentes tipos de sujetos o pacientes (*grupos con diferencias en el pronóstico o con importantes diferencias demográficas, etc.*).

En tales casos, podría requerirse un análisis separado por grupos, para determinar si los resultados son comparables; o, en el peor de los casos, para decidir la exclusión de los individuos detectados como fuente de posibles violaciones al protocolo[14].

Procedimiento de aleatorización

El método para asignar la intervención a los pacientes que participan en el estudio debe estar claramente especificado antes de iniciarse el estudio.

Si el monitor del estudio debe permanecer enmascarado a la secuencia de aleatorización, los códigos de asignación deben ser suministrados a otra persona, que se encargará de preparar los tratamientos[3,21].

El procedimiento puede variar de un experimento a otro, de manera que debe ser lo suficientemente claro para quien tenga la responsabilidad de lograr el consentimiento y hacer la asignación aleatoria.

En términos generales, para cada persona potencialmente elegible se deberá verificar su elegibilidad, y posteriormente, invitarla a participar en el estudio, explicarle en qué consiste y permitirle que conozca y aclare cualquier duda sobre el consentimiento informado. Una vez el paciente firma el consentimiento, se debe proceder a aleatorizar, para lo cual se llamará a la central de aleatorización o se abrirá el sobre sellado que trae la información sobre el tratamiento correspondiente, según como se haya establecido en el protocolo[3,21].

Seguimiento de pacientes durante el estudio

En términos generales, las evaluaciones de seguimiento de los pacientes deben llevarse a cabo en el momento preciso, por parte del personal autorizado para hacerlo y recolectando la información pertinente de acuerdo con la especificación del protocolo.

A pesar de que el grupo se adhiera al protocolo y cumpla a cabalidad las recomendaciones, pueden ocurrir algunas situaciones no anticipadas para las cuales es deseable que se tengan preparados procedimientos de reacción estandarizados (en lo posible, para cada caso particular). En la **tabla 26.8** se pueden apreciar algunas de dichos casos.

Vale la pena mencionar que tanto las pérdidas del seguimiento como cualquier violación al protocolo de estudio deben ser registradas, explicadas y, en lo posible justificadas.

En el caso particular de pacientes que se pierden del seguimiento, se debe intentar obtener información sobre los desenlaces medidos; en particular, si se trata de mortalidad o de morbilidad significativa.

El manejo de eventos adversos y la necesidad de desenmascaramiento de la intervención son situaciones que merecen una mención especial.

Registro de eventos adversos

Para todos los estudios que evalúan intervenciones, es fundamental la estimación de

Tabla 26.8 Problemas no anticipados durante la ejecución del estudio

1. Eventos adversos severos (o la muerte) que requieran el desenmascaramiento de la intervención.
2. Cambios en el personal que maneja a los pacientes del estudio.
3. Modificaciones no aprobadas en el protocolo de estudio.
4. Violaciones al protocolo.
5. Reclutamiento en un número inferior al previsto.
6. Pérdidas del seguimiento.

Fuente: Elaboración propia de las autoras.

seguridad de los tratamientos, de manera que el registro de eventos adversos tiene una importancia capital.

Cualquier cambio indeseable en la salud del sujeto de investigación, asociado o no al uso de un medicamento, y que aparezca durante el desarrollo de un experimento clínico en seres humanos y después de la firma del consentimiento, será considerado un evento adverso, independientemente de que si hay una relación, esta sea o no causal.

Como ejemplos de tales cambios pueden mencionarse: la aparición de una nueva enfermedad (cáncer, infarto cardíaco), la exacerbación de una enfermedad preexistente (crisis de asma), la recurrencia de una enfermedad intermitente (migraña), resultados anormales de laboratorio (hipercolesterolemia, elevación de la creatinina), cambios en los hallazgos del examen físico (aparición de soplos carotideos), accidentes, etc.[6,29,30].

Algunas condiciones no se consideran eventos adversos: por ejemplo, las enfermedades o los síntomas presentes al momento de la inclusión del paciente en el estudio, y que persisten sin cambiar su intensidad o su evolución; el deterioro de la enfermedad bajo estudio, siempre y cuando sea consistente con su progresión típica, y las cirugías programadas por condiciones estables o por otras razones, como una corrección de una hernia que no ha empeorado o cirugías plásticas o reconstructivas programadas.

Una vez identificado el evento adverso, es necesario establecer su severidad, para lo cual se han creado criterios simples.

Criterios de gravedad

- Evento adverso leve: **Síntomas o signos clínicos** fácilmente tolerados por el paciente.
- Evento adverso moderado: **Signos o síntomas** suficientes para interferir con la actividad usual.
- Evento adverso grave: **Incapacidad para trabajar** o desarrollar la actividad usual.

Una vez identificada la gravedad, se debe tratar de establecer la asociación entre el evento adverso y el tratamiento experimental, para lo que también se han desarrollado criterios.

Criterios de imputabilidad[6,20,30]

- Probable: **Relación temporal con la intervención** bajo estudio, sin otros factores causales que expliquen la aparición del efecto adverso; mejoría de los hallazgos con reducción de la dosis o con suspensión de la intervención, y recurrencia ante la nueva exposición, así como la disponibilidad de alguna prueba de laboratorio que identifique la relación causal.
- Posible: **Relación temporal con el tratamiento** experimental, pero en presencia de otros factores causales coexistentes que puedan explicar la aparición del efecto adverso, y ausencia de mejoría con la reducción o la suspensión de la dosis.
- Improbable: **No hay relación temporal, o esta** es dudosa; o bien, hay certeza o probabilidad de otros factores causales.

En ocasiones, cuando aparecen eventos adversos serios, se hace necesario el desenmascaramiento de la intervención.

El procedimiento que se vaya a seguir en tales casos debe haber sido descrito con anticipación y estar detallado en el protocolo.

Reporte de eventos adversos serios

Se considera evento adverso serio cualquier condición nueva que aparece durante el estudio que resulte en ***muerte, amenaza para la vida,*** *hospitalización o prolongación de una ya existente, incapacidad persistente o significativa, o aparición de una anomalía congénita.*

Ante la aparición de un evento adverso grave debe haber claridad en el protocolo acerca de la responsabilidad para reportarlo, formatos específicos y procedimientos para el informe y los tiempos de respuesta.

Si bien puede haber algunas variaciones entre estudios, se han propuesto algunos lineamientos generales sobre cómo proceder en tales casos[6,29,30].

Consideraciones sobre efectos adversos serios:

Deben ser buscados y reportados por las personas en contacto con los sujetos del estudio.

- Deben ser informados inmediatamente, o, al menos, a lo largo de las primeras 24 horas, al monitor del estudio, quien, a su vez, los informará dentro del mismo plazo al comité de seguridad, a la dirección del estudio o al patrocinador, según como sea pertinente.
- Deben ser informados en formularios especiales, con la firma del investigador principal.
- Deben contener, al menos, la siguiente información: fecha y hora de inicio y terminación del evento; clasificación de la severidad; acción tomada en relación con el medicamento de estudio, el tratamiento recibido y el desenlace.

MONITORIZACIÓN EN EXPERIMENTOS CLÍNICOS

La monitorización en este contexto se refiere a todas las acciones diseñadas para supervisar y verificar la planeación, el inicio, la conducción y el procesamiento de datos en los experimentos clínicos.

Objetivos de la monitorización

Hay varias razones por las cuales se hace necesaria la monitorización de los estudios, y, en particular, de los experimentos clínicos[27,31,32], que podrían enmarcarse dentro de dos dimensiones de la actividad de los monitores: 1) una que tiene que ver con aspectos éticos que buscan proteger a los sujetos dentro del marco de una investigación, y 2) otro que tiene que ver con aspectos técnicos mediante los cuales se supervisan las diversas actividades del proyecto[27,33]:

- Evaluación de la adherencia al protocolo:

Un aspecto fundamental durante el desarrollo del estudio es evaluar si los investigadores están siguiendo el protocolo tal como se diseñó.

Inspeccionar la información recolectada para los distintos pacientes permite identificar si existe alguna desviación en el protocolo y facilita informar a los investigadores sobre cualquier falla detectada y sobre la necesidad de adherencia permanente al protocolo. También permite identificar dificultades generales al recolectar la información, que hagan necesario modificarlo[26,31,32].

Debe recordarse que el objetivo de la monitorización es colaborar con el investigador y con su grupo, de manera que deben evitarse el antagonismo y la falta de comunicación.

- Identificación de eventos adversos: La vigilancia activa de los eventos adversos es siempre obligatoria y de la mayor importancia. El proceso de monitorización facilita identificarlos y reportarlos, pues el monitor del estudio actúa como primera línea de comunicación entre el patrocinador del estudio y los investigadores participantes[26,31,32].
- Facilitación del procesamiento de los datos: Es esencial que desde fases muy tempranas de la recolección de información se verifiquen posibles errores, inconsistencias o datos faltantes antes del procesamiento de la información y de su análisis estadístico. Dicho proceso se facilita cuando se planean por anticipado las actividades específicas de monitorización[27,31,32].

Mantener informado al grupo de investigadores

Para preservar el interés entre los investigadores participantes, resulta de gran utilidad informarles sobre el progreso del estudio, las tasas de reclutamiento en los distintos centros involucrados y proveer información sobre los sujetos incluidos.

Los monitores clínicos pueden tener diferentes tipos de entrenamiento y de nivel de estudios, pero es deseable que sean profesionales de la salud, para un mejor entendimiento de la terminología utilizada, de los procedimientos y de los diversos formatos para recolectar la información[8,13,32].

De acuerdo con su dependencia, los monitores pueden ser:

Monitores internos o coordinadores del estudio

Dependen directamente del grupo de investigadores en un sitio en particular, y, por lo general, desarrollan la función de seguimiento del proceso de recolección de la información[8,13,32].

Monitores externos

Designados por los patrocinadores del estudio, por organizaciones de investigación contratada o por agencias reguladoras gubernamentales, cumplen la función de asegurar que la conducción del experimento clínico, la recolección de la información y el reporte de los datos se ajusten al protocolo y se ciñan estrictamente a las normas de buenas prácticas clínicas[32].

Funciones del monitor del estudio

Según la guía de buenas prácticas clínicas, es responsabilidad del monitor llevar a cabo las siguientes actividades para garantizar que el experimento clínico sea adecuadamente conducido y documentado[2,31-33]:

- Actuar como principal puente de comunicación entre el patrocinador y el investigador. La comunicación es el elemento fundamental para la adecuada ejecución del estudio. En el proceso de contactar el sitio de estudio y al investigador, es necesario proveer una gran cantidad de información para conducir la investigación, pero este volumen de documentos no necesariamente significa comprender y apropiar los pasos por seguir durante la ejecución. El papel del monitor en este proceso es de suma importancia, y debe lograr empatía con el investigador, para convertirse en miembro imprescindible del equipo del estudio, que actúa como entrenador[27,33].
- Verificar que el investigador está adecuadamente calificado y cuenta con los recursos necesarios para el desarrollo del estudio, tanto al comienzo de la investigación como durante toda su ejecución. En este aspecto, el papel actual del monitor va más allá de comprobar las condiciones del investigador y del sitio para el estudio; debe convertirse, además, en un puente de información entre el patrocinador y el investigador clínico, facilitar que los aportes de este último sean tenidos en cuenta y, eventualmente, generen ajustes en el protocolo, los cuales permitirán conducir mejor el estudio y anticipar posibles dificultades que se generan en contextos particulares[27,33].
- Confirmar, en relación con el producto que está siendo investigado:
  - Que las condiciones y tiempo de almacenamiento son aceptables, y que hay disponibilidad suficiente del producto a lo largo de todo el estudio.
  - Que el producto bajo investigación está siendo administrado únicamente a individuos elegibles, según el protocolo de estudio y según las dosis especificadas por dicho estudio.
  - Que los sujetos participantes en el estudio reciben la instrucción suficiente acerca del uso, el manejo, el almacenamiento y la devolución del producto bajo investigación.
  - Que el recibo, el uso y la devolución del producto de investigación están adecuadamente controlados y documentados en los centros participantes en el estudio.
  - Que el destino final del producto bajo investigación que no ha sido utilizado en los centros participantes en el experimento se está llevando a cabo según los procedimientos autorizados por el patrocinador, y que cumple con los requerimientos de agencias reguladoras.



- Que el investigador sigue estrictamente el protocolo de estudio y sus modificaciones, debidamente aprobadas.
- Que el consentimiento informado ha sido firmado por cada sujeto antes de iniciar su participación en el estudio.

- Que el investigador ha recibido todos los documentos y los suministros necesarios para conducir adecuadamente el experimento.
- Que el investigador y todo su equipo de trabajo están suficientemente informados sobre el protocolo de estudio y sobre cómo se llevará a cabo el experimento.
- Que el investigador y su equipo de trabajo están desempeñando personalmente las funciones especificadas para el estudio de acuerdo con el protocolo, y que no han delegado dichas funciones en personal no autorizado.

- Que el investigador está incluyendo en el estudio solo a sujetos elegibles de acuerdo con los criterios de inclusión y de exclusión descritos en el protocolo.
- Que se reporta la tasa de reclutamiento de los sujetos en el estudio.

- Que los documentos fuente (historia clínica, reportes oficiales de laboratorio, registros electrofisiológicos, electrocardiogramas, radiografías, tomografías, etc.) y otros formatos de registro del experimento son confiables, veraces, completos y legibles, y que están al día y archivados adecuadamente.

- Que la documentación necesaria para el inicio, la conducción y la culminación del experimento es veraz, completa, oportuna y legible, y está fechada e identificada con el nombre del estudio.
- Que se hace un diligenciamiento exacto y completo de los formatos y los documentos del estudio, y que se comparan entre sí. El monitor, específicamente, debe verificar que:
  - Los datos requeridos por el protocolo están siendo reportados de manera confiable en el formato de reporte de caso (**CRF, por las** siglas en inglés de ***case report form***) **y son** consistentes con la información consignada en los documentos fuente.
  - La dosificación o cualquier modificación de ella está adecuadamente documentada para cada paciente en el experimento.
  - Los eventos adversos, las medicaciones concomitantes o las enfermedades intercurrentes son reportados, de acuerdo con el protocolo, en el **CRF.**
  - Las visitas a las que los pacientes fallan, las pruebas diagnósticas y los exámenes físicos que no se llevan a cabo son registrados como tal en los **CRF.**
  - Todos los retiros y las pérdidas del seguimiento de individuos participantes en el estudio son reportados y explicados en el **CRF.**
  - Se informa al investigador sobre cualquier error, cualquier omisión o cualquier ilegibilidad en el **CRF. El monitor debe asegu**rarse de que se hacen las correcciones, las eliminaciones o las adiciones apropiadas, y que, adicionalmente, se identifica con fecha, iniciales del investigador y explicación (si es necesario) cualquier modificación en este formato.
  - Todos los eventos adversos han sido adecuadamente reportados dentro del período requerido por las Guías de Buenas Prácticas Clínicas, el protocolo del estudio, el comité de ética y el patrocinador.
- Que el investigador mantiene un archivo con los documentos esenciales para la conducción de un experimento clínico.
- Que se identifica y se comunica al investigador sobre cualquier desviación del protocolo, de las buenas prácticas clínicas o de los requerimientos de las entidades reguladoras, y que se toman las acciones apropiadas diseñadas para prevenir la recurrencia de las desviaciones detectadas.

Estrategias de monitorización

Es posible monitorizar por diversos métodos el desarrollo del proceso de recolectar información para un estudio clínico; entre ellos, se encuentran: llamadas telefónicas, correo electrónico, fax, cartas, análisis de registros electrónicos del estudio, visitas o una combinación de todas las anteriores[27,32,34].

Aunque no existe acuerdo sobre cuál puede ser el mejor método, es posible llevarse sorpresas en lo referente a la confiabilidad de la información cuando no se tiene un control directo sobre los datos que se están recolectando, por lo cual debe haber al menos una estrategia que permita la verificación personal de cada registro contra los documentos fuente (historias clínicas, reportes originales de laboratorio, electrocardiogramas, etc.).

Por lo anterior, es recomendable que se establezca ***a priori*** **algún mecanismo de control** sobre cualquier método de monitorización que no implique visitas directas a los centros participantes en el estudio[32].

Sobre la frecuencia aproximada que deben tener las visitas de monitorización, no existe tampoco un número "mágico" o absoluto, y depende de varios factores[32]:

- El grado de prioridad del estudio clínico para el patrocinador.
- La complejidad del protocolo del estudio.
- La tasa de reclutamiento de los sujetos o de los pacientes.

- El tiempo transcurrido desde la visita anterior.
- Los problemas observados por el monitor durante visitas previas.
- Los problemas reportados por los centros participantes desde la visita anterior.

Una vez considerados todos esos factores, se debe establecer la frecuencia deseable de las visitas de monitoreo antes del inicio del estudio, pero permitiendo cierta flexibilidad a la intensidad de estas durante el desarrollo de la investigación, de acuerdo con las necesidades que se vayan presentando[32].

Para estudios clínicos con doble o triple cegamiento, es de vital importancia que el monitor también permanezca cegado al grupo de intervención al cual pertenece cada sujeto aleatorizado, con el fin de evitar la posibilidad de introducir sesgos de medición durante el desarrollo del experimento[32].

Visitas de monitorización

Antes de cada visita, el investigador o el coordinador del estudio en cada centro deben ser contactados para que conozcan el objetivo de la visita, estén preparados y calculen el tiempo necesario para su realización. Adicionalmente, deben preparar lo siguiente[31,32]:

- Los **CRF actualizados y con las correcciones** que hayan quedado pendientes de la visita anterior.
- Una lista de los sujetos tamizados para el estudio y el número de los incluidos definitivamente, así como cualquier otro punto que justifique discusión.
- Un sitio para que el monitor revise toda la información durante el desarrollo de la visita.

Es muy importante que el monitor sea visto como un colega y un miembro del grupo de trabajo, y no como un vigilante que solo busca detectar los errores cometidos durante su ausencia, pues una adecuada interacción favorece el proceso de recolectar información y garantiza una óptima conducción de la investigación[31,32]. Y es también responsabilidad de quien lo entrena, y del monitor, que el comportamiento y la intención sean los de ser un colega y estén del mismo lado.

Durante la visita de monitorización, el investigador principal y sus coinvestigadores (si es posible), así como el coordinador del estudio (si lo hay), deben estar disponibles.

Sin embargo, antes de reunirse con ellos, el monitor debe tener tiempo suficiente para revisar todos los CRF **y los respectivos documentos fuente (historias** clínicas, reportes originales de laboratorio, etc.), de manera independiente, para detectar omisiones y verificar la adherencia al protocolo, la legibilidad y la consistencia de la información reportada, de tal forma que pueda tener una reunión productiva con los miembros del equipo de trabajo[31,32].

El monitor puede revisar durante la visita la cantidad de medicamentos suministrados y devueltos; si ha habido eventos adversos severos o serios, puede ser necesario que tenga a mano información adicional diferente de la historia clínica del paciente. Sin embargo, cuando esto suceda, se le debe advertir al investigador, bien sea telefónicamente o por fax, sobre cuáles documentos adicionales serán necesarios para su revisión[19].

Después de la visita, es importante que el investigador principal y su grupo de trabajo devuelvan los documentos fuente a los sitios de origen del centro participante, si fuera el caso, y que todas las correcciones y el diligenciamiento completo de la información faltante en los **CRF se lleven a cabo** tan pronto como sea posible, para evitar nuevas omisiones y errores adicionales[32].

Es responsabilidad del patrocinador del estudio o del grupo investigador diseñar los formatos que serán diligenciados por el monitor durante sus visitas a los centros participantes. Estos formatos deben contener toda la información pertinente relacionada con la conducción de la investigación[32].

El plan de monitoreo y la decisión sobre la información que se debe incluir en esos registros puede variar según los objetivos de la investigación y del patrocinador, pero, en términos generales, constituyen listas de chequeo que recolectan o recuerdan considerar la siguiente información, dependiendo del momento del desarrollo del estudio[32]:

Durante la preparación del estudio-inicio de la investigación

- Calificación del investigador en cada centro participante en el estudio: el investigador

debe demostrar un nivel apropiado de educación, experiencia y entrenamiento para conducir la investigación a través de una hoja de vida actualizada.
- Disponibilidad de tiempo para el estudio por parte del investigador.
- Completo conocimiento del protocolo por parte del investigador y de los delegados suyos participantes en el estudio.
- Conocimiento, por parte del investigador, de los CRF, del consentimiento informado, de los formatos de reporte de eventos adversos, del manejo de los medicamentos suministrados para el estudio, de la disponibilidad de documentos fuente para su revisión, del mantenimiento de los registros de cada paciente y de las inspecciones por parte de entidades reguladoras.
- Demostrar capacidad para reclutar al número de sujetos necesarios en el tiempo propuesto por el patrocinador.
- Documentar la presentación del protocolo al comité de ética para su aprobación, e incluir para ello el consentimiento informado.

Durante la inclusión de pacientes

- Revisar la visita de monitorización anterior y resolver los problemas que habían quedado pendientes desde entonces.
- Revisar los CRF y los documentos fuente (historia clínica, reportes originales de laboratorio, electrocardiogramas, radiografías, tomografías, etc.) buscando la aparición de eventos adversos serios, para confirmar que estos han sido reportados de manera oportuna y en el formato correspondiente.
- Monitorizar el estado del reclutamiento de los pacientes hasta el momento de la visita del monitor.
- Revisar los formatos de consentimiento informado.
- Revisar los CRF en cuanto a su legibilidad, su consistencia, su adherencia al protocolo de estudio, su exactitud y su completo diligenciamiento.
- Verificar la información registrada en los formatos de CRF contra la de los documentos fuente.
- Monitorizar los recursos del centro para proseguir con la investigación y el suministro adecuado de medicamentos de estudio.

Durante la finalización del estudio-visitas de cierre

- Confirmar que los requerimientos de las buenas prácticas clínicas fueron seguidos estrictamente en lo relacionado con la firma del consentimiento informado.
- Revisar los CRF pendientes, para confirmar su completo diligenciamiento, y hacer la respectiva verificación contra los documentos fuente.
- Confirmar que las correcciones de inconsistencias en los CRF se llevaron a cabo de manera adecuada.
- Verificar que el conteo del medicamento de estudio se hizo oportunamente y de forma confiable.
- Asegurarse de la disposición final y de la devolución de los medicamentos no utilizados durante el estudio.
- Confirmar la devolución al patrocinador de todos los suministros diferentes de la medicación del estudio.
- Explicar la obligación de archivar los registros de pacientes incluidos en el centro participante.
- Asegurarse de que el reporte de todos los eventos adversos serios fue completo, y de que se garantizó un seguimiento adecuado y suficiente a las personas que los presentaron.
- Recordarle al centro participante la necesidad de informar al comité de ética que el estudio concluyó.
- Inspeccionar el archivo de todos los documentos del estudio que debe conservar el investigador, y verificar que no falte ningún registro.

ESTRATEGIAS PARA MEJORAR EL DESARROLLO DEL ESTUDIO

El papel de los monitores durante la planeación y la conducción de un experimento clínico es muy activo en la actualidad, y a menudo participan en la selección de los investigadores y de los centros participantes, en el diseño de los CRF, en la propuesta de modificaciones al protocolo de estudio, en el seguimiento de la recolección de información, en el reporte de eventos adversos y hasta en la verificación de datos durante la fase de análisis de resultados[32-34].

La participación del monitor en la selección de investigadores debe basarse en verificar su competencia y su capacidad para conducir el estudio de manera eficiente y en perfecta adherencia a las buenas prácticas clínicas.



Se espera, adicionalmente, que el investigador seleccionado mantenga una comunicación abierta y oportuna, así como un sentido de colaboración permanente, con el monitor y con el patrocinador[32,33].

Durante el desarrollo del protocolo, el monitor también puede formar parte del equipo de trabajo, y, gracias a su comunicación con el patrocinador y los investigadores, puede sugerir modificaciones basadas en su experiencia previa y en su conocimiento práctico y su habilidad.

El monitor debe ser tratado como parte integral del equipo de trabajo por el investigador y sus colaboradores, y la relación con ellos debe basarse en el mutuo respeto. La opinión del monitor debe ser escuchada y evaluada para considerar sus recomendaciones elemento fundamental en el aseguramiento continuo de calidad durante la ejecución del estudio. Adicionalmente, el monitor puede ejercer presión sobre el investigador si su desempeño en la conducción de la investigación es inadecuado y si no está cumpliendo con los compromisos adquiridos en un comienzo[32].

El monitor puede percibir que la calidad de la participación por parte del investigador ha variado y se ha deteriorado a lo largo de la ejecución del estudio. Los problemas identificados más a menudo cuando esto sucede son los siguientes[32]:

- Las garantías y los incentivos ofrecidos a los pacientes en el estudio resultan insuficientes para que un número adecuado de sujetos termine siendo aleatorizado.
- El protocolo exige demasiado tiempo y esfuerzo por parte de los potenciales participantes, que terminan abandonando el estudio.
- El protocolo exige demasiado tiempo y esfuerzo para el investigador y su grupo de trabajo.
- El número de personas en el equipo de trabajo del investigador es insuficiente, o su entrenamiento es inadecuado para el seguimiento de los pacientes.
- Las citas para las visitas de seguimiento son programadas en fechas, horas o días que resultan inconvenientes para los participantes.

Ante un funcionamiento inadecuado por parte de un centro o de un investigador participante, después de intentar identificar el problema y buscar las razones que pueden estar contribuyendo al deterioro de la calidad en la conducción del estudio, es fundamental que, mediante comunicación abierta entre el monitor y el investigador, se busquen posibles soluciones al problema y se implementen de manera inmediata[32].

FINALIZACIÓN DEL ESTUDIO-ACTIVIDADES DE CIERRE

Hay varias razones por las cuales se puede terminar un estudio; algunas son positivas, como la terminación exitosa de un proyecto que logró los objetivos propuestos, o la terminación anticipada porque se haya establecido que la terapia en experimentación tiene un efecto benéfico tal que no solo no es necesario continuar, sino que al hacerlo se estaría faltando a la ética, pues ya se encontró una respuesta clara y se estaría negando el beneficio del tratamiento a un grupo.

Algunas de las razones pueden, sin embargo, ser negativas: bien sea porque la terapia en experimentación ha fallado, ha demostrado no ser eficiente o causa más daño que beneficio, o porque el proyecto de investigación por sí mismo no está alcanzando los objetivos propuestos. Entre las causas para esto último puede haber fallas en enrolar al número adecuado de pacientes, fallas del grupo investigador o de sus colaboradores en mantener adecuadamente los estándares de calidad de la investigación o falta de adherencia al protocolo.

Para definir cuándo terminar un proyecto que no está siendo exitoso, es muy importante tener una visión objetiva de la situación y considerar tanto los logros alcanzados como los fracasos que se presentan. Especialmente en los estudios experimentales para probar nuevos medicamentos, hay presiones tanto del patrocinador como de interés del investigador para mantener vivo el estudio, así sea con una perspectiva poco realista. La objetividad es la única herramienta necesaria para terminar a tiempo un estudio que falle en cualquiera de los aspectos mencionados.

Se han definido algunas consideraciones para la terminación de un estudio que no esté logrando los objetivos o del cual se prevea que no los logrará. Los recursos pueden ser utilizados en otra investigación más factible, el grupo de colaboradores no se verá decepcionado ante un fracaso si el estudio se detiene a tiempo, se pueden evitar mayores costos futuros y las enseñanzas que deja la experiencia del proyecto terminado pueden beneficiar otros proyectos en ejecución; en especial, si están siendo adelantados en la misma área de investigación.



Establecer con anterioridad pautas claras que permitan definir cuándo terminar el proyecto facilita la toma de decisiones en el momento en el que se presenten dificultades durante su desarrollo.

Por tanto, en el protocolo debe quedar establecido cuándo se considerará que la terapia experimental está fallando (falla terapéutica), en el caso de estudios experimentales, o cuándo los objetivos de la investigación no se cumplen por otras razones, como una baja tasa de reclutamiento, por una selección inadecuada de la población de estudio o por fallas en los métodos planteados.

Al identificarse fallas en el protocolo de investigación, es posible hacer enmiendas a este antes de dar por terminado el proyecto. Los investigadores deben estar conscientes de que cualquier enmienda requiere someter nuevamente el protocolo a la evaluación de los comités reguladores correspondientes antes de ser adoptada como parte del estudio.

Algunas de las actividades que deben cumplirse antes de dar por terminado un proyecto incluyen:

a) **Revisión de los datos por parte de un comité** externo previamente definido, con el fin de obtener sugerencias sobre alternativas de solución.
b) **Redefinición de los objetivos o redirección** del proyecto.
c) **Identificación de un nuevo líder o coordinador** del proyecto cuando se identifica que la falla reside en la forma como está siendo conducido.
d) **Consulta al grupo de colaboradores sobre** posibles soluciones que permitan revivir el proyecto. Si se anticipa que una enmienda al protocolo no podrá mejorar su factibilidad, se debe pensar en la mejor forma de darlo por terminado.

ACTIVIDADES DE CIERRE DE UN PROYECTO

Recolección y archivo del total de la información (datos) obtenida

En caso de estudios multicéntricos o multisitio, el monitor debe asegurarse de que todos los centros participantes han entregado las formas completas de recolección de datos de los pacientes ingresados al estudio, bien al centro coordinador, o bien, al de manejo de datos. El coordinador del estudio debe cerciorarse de que el número de pacientes reportados en los resultados del análisis estadístico corresponde al reportado por los asistentes de recolección de los diferentes centros participantes.

Una vez terminado el análisis de la información y verificados sus resultados, se puede proceder al archivo de los formatos de recolección, que deben conservarse por dos años después de finalizada la investigación, y hasta por cinco años en el caso de experimentos para el desarrollo de medicamentos. Hay diferentes métodos para hacer este archivo, como el archivo de las formas originales de forma clasificada por centro investigador, por orden secuencial de acuerdo con el número de identificación de cada paciente en el estudio, o por archivos electrónicos de microfilmación de las formas originales de recolección, que tienen la ventaja de requerir menor cantidad de espacio físico para su conservación.

Recolección de los equipos e insumos entregados para hacer la investigación

Cada investigador determinará qué hacer con los insumos entregados en estudios multicéntricos a los diferentes centros de investigación. Es de particular relevancia en experimentos clínicos para evaluación de medicamentos encargar a una persona de destruir el medicamento sobrante y de llevar un registro del número de frascos o de ampollas, así como de la cantidad de medicamento desechado y del día. Esto ayudará a tener un control de dicha actividad.

Elaboración y entrega de informes

Es usual que en estudios patrocinados por cualquier entidad se deba entregar un informe detallado de las actividades llevadas a cabo y de los resultados obtenidos. Si el estudio no tiene patrocinador, por lo general, el ente financiador requiere que se le informe de manera detallada sobre el uso de los recursos dados para la ejecución del proyecto. Esta tarea puede ser dispendiosa, por lo cual desde el comienzo del estudio se deben plantear estrategias de control de los recursos financieros, y se debe identificar a una persona idónea para que lleve la contabilidad del proyecto.

TENDENCIAS ACTUALES EN LA MONITORIZACIÓN DE ESTUDIOS CLÍNICOS

Es necesario no perder de vista el propósito de la monitorización, la cual busca proteger a los sujetos bajo investigación, asegurar la calidad de los datos y cumplir con los requerimientos de las agencias regulatorias según el contexto.

Se ha visto en el pasado una tendencia a ejercer esta actividad como un intenso análisis retrospectivo de los documentos que demuestran el proceso de ejecución de un estudio clínico, sin haber jugado un papel más preventivo en su interacción con el investigador y con los sitios de estudio[33,34].

Recientemente, se ha introducido el concepto de ***monitorización basada en riesgo***, que busca identificar áreas particulares del estudio susceptibles a desviaciones del protocolo o a errores sistemáticos. Dicha aproximación pretende reducir las visitas y la verificación contra documentos fuente, y se concentra prioritariamente en aspectos de riesgo y focaliza los esfuerzos de revisión en torno a estos.

Hoy día se espera un papel más activo del monitor en su interacción con los sitios de investigación, con el equipo que participa en el estudio y con el investigador clínico, que permita identificar procesos con mayor probabilidad de falla, para intensificar, de manera preventiva, los esfuerzos de entrenamiento del equipo, y así reducir el riesgo de problemas durante la conducción del estudio[33-36].

CONCLUSIONES

Para la culminación exitosa de cualquier estudio clínico se requiere no solo una conducción adecuada, sino una buena monitorización.

Algunos pasos de la investigación clínica son críticos para la calidad y la aceptabilidad de los resultados. La conducción y la monitorización adecuadas garantizan que se hace una permanente evaluación de la adherencia al protocolo por parte del investigador y de su grupo de trabajo; que se identificarán oportunamente los eventos adversos; que la recolección y el procesamiento de la información serán óptimos, y que se mantendrá el espíritu de trabajo en el grupo de investigadores.

La conducción del estudio se facilita mediante el uso de herramientas que permitan un seguimiento de las actividades en cada una de las fases del estudio, para lo cual se recomiendan formatos de registro de seguimiento de datos, informes de recolección etc.

De igual forma, durante el monitoreo de la recolección de información en cada uno de los centros participantes en el estudio, es necesario llevar un registro de las visitas y del seguimiento que se hagan, en los formatos diseñados para tal fin. Estos formatos no son otra cosa sino listas de verificación que el monitor debe completar en cada una de sus visitas.

El papel del coordinador es cada vez más activo durante el diseño, la conducción y el análisis de los datos. Su opinión debe ser escuchada, pues sus recomendaciones buscan corregir los errores detectados durante el desarrollo del estudio, vigilar el correcto diligenciamiento de los CRF, y garantizar la información oportuna y completa de eventos adversos.

Referencias

1. Bangdiwala S. Quality Assurance Methods in Multicentre Clinical Trials. Workshop Materials. Kunming. 2003.
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Guideline for Good

Clinical Practice. Rockville, MD: ICH; 1996.
3. Knatterud GL. Management and conduct of randomized controlled trials. Epidemiol Rev. 2002; 24:12-25.
4. Spilker B. Standardizing information across protocols. En: Spilker B, editor. Guide to clinical trials. 2nd ed. Philadelphia: Lippincott Raven; 1996. p. 225-34.
5. Spilker B. Choosing clinical investigators. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 410-7.
6. Califf RM, Morse MA, Wittes J, et al. Toward protecting the safety of participants in clinical trials. Control Clin Trials. 2003;24:256-71.
7. Spilker B. Data review committees. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 124-8.
8. Hulley SB, Cummings SR. Diseño de la investigación clínica. Madrid: Doyma; 1997.
9. Lovato LC, Hill K, Hertert S, et al. Recruitment for controlled clinical trials: literature summary and annotated bibliography. Control Clin Trials. 1997;18:328-52.
10. Rand CS, Sevick MA. Ethics in adherence promotion and monitoring. Control Clin Trials. 2000;21(5 Suppl):241S-7S.
11. Spilker B. Ethical considerations and issues. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 245-57.
12. Colombia, Ministerio de la Protección Social (MPS). Resolución 002378 por la cual se adoptan las Buenas Prácticas Clínicas para las instituciones que conducen investigación con medicamentos en seres humanos. Bogotá: MPS, Diario Oficial 47033; 2008.
13. Pocock SJ. Monitoring trial progress. En: Pocock SJ, editor. Clinical trials, a practical approach. Chichester: John Wiley & Sons; 1983. p. 142-59.
14. Spilker B. Clinical trial initiation and conduct. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 423-9.
15. Bourque LB, Clark VA. Processing data: The survey example. Newbury Park: Sage Publications; 1992.
16. Hernández R, Fernández C, Baptista P. Metodología de la investigación. México: McGraw Hill; 1998.
17. Spilker B, Schoenfelder J. Data collection forms in clinical trials. New York: Raven Press; 1991.
18. Spilker B. Preparing data collection forms. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 262-71.
19. Pérez A, Cáceres CA. El sistema de manejo de datos en los estudios de investigación: Métodos y guías. Rev Colomb Neumol. 1999;11:126-33.
20. Friedman LM. Data collection forms. En: Friedman F, editor. Fundamentals of clinical trials. New York: Springer; 1998. p. 72-89.
21. Friedman LM. The randomisation process. Fundamentals of clinical trials. Littleton, MA: PSG Publishing Co; 1985. p. 51-9.
22. Spilker B. Preparing, packaging, and dispensing of clinical trial medications. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 188-95.
23. Stouthamer-Loeber M. Data collection and management: a practical guide. Newbury Park: Sage Publications; 1998.
24. Spilker B. Training professionals and their staff in clinical trial methods. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 382-9.
25. Gore SM. Assessing clinical trials- first steps. Br Med J (Clin Res Ed). 1981;282:1605-7.
26. Knatterud GL, Rockhold FW, George SL, et al. Guidelines for quality assurance in multicenter trials: a position paper. Control Clin Trials. 1998;19:477-93.
27. Gomez-Panzani E. Clinical study conduct and monitoring. En: Chin RY, Bairu M, editors. Global clinical trials: effective implementation and management. London: Elsevier; 2011. p. 423-69.
28. Weigle K, Porter C, Rodríguez N. Notas de clase, curso corto en métodos de conducción de estudios. 1995.
29. Spilker B. Systems for classifying diseases and adverse reactions. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 479-81.
30. Spilker B. Interpretation of adverse reactions. En: Spilker B, editor. Guide to clinical trials. Philadelphia: Lippincott Raven; 1996. p. 565-87.
31. Kolman J, Meng P, Scott G. Good clinical practice-standard operating procedures for clinical researchers. Chichester: John Wiley & Sons; 1998.
32. Spilker B. Monitoring a clinical trial. En: Spilker B, editor. Guide to clinical trials. 2nd ed. Philadelphia: Lippincott Raven; 1996. p. 430-48.
33. Ansmann EB, Hecht A, Henn DK, et al. The future of monitoring in clinical research - a holistic approach: linking risk-based monitoring with quality management principles. Ger Med Sci GMS E-J. 2013;11:Doc04.
34. Romero GAS. Clinical research monitoring: scenarios and challenges. Rev Soc Bras Med Trop. 2013;46:3-6.
35. Mitchel J, Kim Y , Hamrell M, et al. Three-pronged approach to optimizing trial monitoring. Applied Clinical Trials. 2014;23:37-44.
36. Wang A, Clarke KA, Little J, et al. Shifting monitoring paradigms: challenges and opportunities. Applied Clinical Trials. 2014;30-6.

CAPÍTULO

27

Adherencia

Álvaro J. Ruiz Morales
Claudia Marcela Granados Rugeles

En un experimento clínico reciente se demostró que en uno de los brazos de tratamiento de la insuficiencia cardíaca, la mortalidad fue 34 % menor en quienes tuvieron mejor cumplimiento ¡con el placebo![1]

INTRODUCCIÓN

La insuficiencia cardíaca congestiva está asociada a una alta mortalidad, y las intervenciones buscan reducirla y, a la vez, mejorar la calidad de vida de los pacientes. El cumplimiento con la intervención (toma de, al menos, el 80 % de las dosis del medicamento) se asoció, a su vez, a una mortalidad el 35 % menor que en quienes no la cumplieron. El tratamiento activo con candesartán, mostró una reducción significativa en la mortalidad (10 %; $p = 0{,}032$), en comparación con el placebo. Pero el cumplimiento con la intervención asignada, fuera activa o placebo mostró una reducción del 35 % de la mortalidad.

El resultado del estudio CHARM no fue una sorpresa para los lectores. Ya previamente se habían encontrado resultados similares en experimentos clínicos, con intervenciones con bloqueadores β en infartos de miocardio[2], con pacientes antiarrítmicos en arritmias postinfarto[3], y con hipolipemiantes en postinfarto[4], en quienes el grupo con la mejor adherencia tuvo mejores resultados. Una explicación lógica para estos resultados puede ser que los pacientes más enfermos tienen más dificultades para ser adherentes, pero en dichos estudios y otros similares, los cuales han tenido resultados que apuntan en la misma dirección, no se ha encontrado que la severidad sea la determinante del peor pronóstico[5].

Una explicación más aceptable para el resultado es que el cumplimiento con la intervención farmacológica está más probablemente asociada a una mejor adherencia (mayor cumplimiento también con las otras recomendaciones)[6], como control del peso, de la ingesta de sodio y de la actividad física y el cuidado con la alimentación, y con el control de los otros factores de riesgo, que también han demostrado reducir la morbilidad y la mortalidad[7].

Muchos estudios han demostrado que la adherencia a las recomendaciones diferentes de la intervención farmacológica tiene efectos significativos. Y han mostrado cómo el paciente que no cumple con la medicación es menos adherente a las otras intervenciones[8] y tiene desenlaces menos satisfactorios[9].

Por todo lo anterior, tiene gran importancia conocer los determinantes de la adherencia, para prevenir o para hacer modificaciones de refuerzo en acciones que busquen mejorar los desenlaces. Si se planea un estudio para evaluar, por ejemplo, la efectividad de un medicamento antituberculoso en pacientes con tuberculosis resistente, hay muchas preguntas relevantes que se deben resolver. ¿Cómo asegurarse de que el paciente va a seguir sus instrucciones? ¿Cuáles

son las intervenciones que han mostrado tener los mejores resultados, y cuáles, las que han demostrado tener mejor adherencia? ¿Cuáles acciones tienen impacto en el cumplimiento de las recomendaciones? ¿Cuáles son las particularidades que diferencian el seguimiento de las instrucciones cuando se hace parte de un estudio vigilado del seguimiento de estas en la vida real, en un paciente que no pertenece a un estudio? ¿Hay diferencias cuando las instrucciones deben ser seguidas por tiempos prolongados (meses o años) y cuando son para un ciclo corto de días?

Se han descrito muchos factores asociados a la probabilidad de adherencia, relacionados con el paciente, con la intervención, con la condición que se trata, con el ambiente[10] en el que vive el individuo, con las explicaciones y la educación que reciba este del médico, con la complejidad de las recomendaciones, y muchos más[11].

ADHERENCIA, CUMPLIMIENTO Y CONCORDANCIA

La adherencia es el grado en el que el comportamiento de una persona es consistente con las recomendaciones impartidas para el cuidado de su salud[12]. O el grado en el que los pacientes siguen las instrucciones recibidas.

No se trata solamente de que se sigan las indicaciones sobre los medicamentos, sino también las recomendaciones no farmacológicas, las cuales, a su vez, pueden incluir cambios en la alimentación, actividad física, control del peso, abandono del tabaquismo y cumplimiento con otras instrucciones para pruebas diagnósticas o para terapias adicionales, por parte de diferentes profesionales de la salud[13]. La decisión de la persona tiene varios determinantes, relacionados, a su vez, con el médico, con la propia situación de salud y con sus percepciones, sus creencias, su comprensión del problema y su actitud general.

El concepto de adherencia es amplio, e incluye el cumplimiento con las recomendaciones farmacológicas y con las no farmacológicas, pero debe complementarse con dos aspectos adicionales: la ***concordancia*** y la *persistencia*.

Se habla de concordancia como el necesario acuerdo entre el médico y el paciente, lo cual implica que se entienda la importancia de la enfermedad, de las intervenciones propuestas, las consecuencias del incumplimiento y el compromiso de seguimiento por parte del paciente[13]. Y la persistencia se refiere al seguimiento de las indicaciones por un tiempo recomendado[14]. Ambos conceptos deben cumplirse para lograr la adherencia[15].

FACTORES QUE INFLUYEN EN LA ADHERENCIA

La adherencia tiene múltiples componentes, y debe ser considerada un comportamiento complejo, que depende de múltiples instancias. Hay factores que pueden facilitar o aumentar la adherencia, y también, factores que pueden presentarse como barreras que la dificultan. Algunos pueden modificarse, y otros no, pero es importante identificarlos, pues debe intentarse su prevención o su modificación; y cuando no sea posible ninguna de estas dos opciones, deben reforzarse otras conductas que faciliten el tratamiento.

Las causas de no adherencia y de incumplimiento pueden clasificarse en seis grandes categorías, que se resumen en la **tabla 27.1**.

Factores socioeconómicos

Son los primeros que se ofrecen como explicación cuando un paciente incumple con su tratamiento, aunque no son su causa más frecuente. Si bien las dificultades económicas pueden originar incumplimiento, otras razones, como se verá más adelante, tienen más fuerza como explicación.

Puede haber dificultades por razones económicas: por ***competencia de prioridades***, cuando los recursos, ya de por sí escasos, deben ser distribuidos según diferentes urgencias. A veces, el sustento diario y los gastos elementales de vivienda y transporte son más importantes (o, al menos, más urgentes) para una familia que el gasto en salud. Pero también puede ocurrir que haya una ***falta absoluta de recursos*** para adquirir los medicamentos o para los cambios de estilo de vida sugeridos. En ninguna de tales situaciones la solución es fácil, ni está al alcance del médico o del equipo de salud; y aunque siem-

Tabla 27.1 Factores que influyen en la no adherencia o en el incumplimiento

• Socioeconómicos
• Dependientes de la enfermedad
• Relacionados con el paciente
• Del médico y del equipo de salud
• Originados en la intervención
• Provenientes del sistema de salud

pre debe hacerse la evaluación crítica de lo que se recomienda, en busca de las intervenciones más eficientes, en estos casos esa revisión debe hacerse con particular cautela.

Otras causas que pueden estar relacionadas con bajos recursos, pero sobre las que se puede influir más fácilmente son el ***bajo nivel educativo***, que dificulta comprender los conceptos, incorporar cambios en el estilo de vida o apropiar el conocimiento sobre la enfermedad y sobre la necesidad del tratamiento. También puede influir la ***falta de redes de apoyo***[16]; sobre todo, en pacientes con disfunciones sensoriales, con limitaciones físicas o con alteraciones cognoscitivas. Es más difícil para un paciente que tiene reducida su visión cumplir con las medicaciones, y más difícil, para quien dependa de un cuidador para recibir los medicamentos, o para preparar los alimentos. Finalmente, la *cultura y las creencias populares* son influencias que se deben considerar. En algunas culturas, por ejemplo, aun si no tomar el medicamento pone en peligro la vida, se prefiere este riesgo a presionar a un anciano, porque puede parecer irrespetuoso[17]; o puede verse la enfermedad como un castigo que debe aceptarse, o como un método de purificación[18].

Factores relacionados con la enfermedad

La gravedad de la enfermedad puede ser un estímulo poderoso para tomar el medicamento, o para cumplir con las recomendaciones. Si bien se esperaría que a mayor gravedad de la enfermedad haya mejor adherencia, algunos estudios[19] han mostrado que no solo no es el caso, sino que en el extremo de mayor gravedad la adherencia es ***menor***. Los resultados son variables[20], y la conclusión de un metaanálisis es que la adherencia es mayor: 1) cuando se han prescrito medicamentos (más que recomendaciones no farmacológicas), 2) en los adultos que en los niños, y 3) en la infección por el virus de la inmunodeficiencia humana (VIH), la artritis, las enfermedades gastrointestinales y el cáncer (Ca); y significativamente menor, en la diabetes, las enfermedades pulmonares y las alteraciones de sueño.

La educación es importante, y la utilidad percibida de la intervención juega un papel primordial en la adherencia. Si el paciente entiende la utilidad de lo prescrito, tiene mayores posibilidades de permanecer adherente[21]. Algunas enfermedades han sido señaladas como causantes de menor adherencia; específicamente, la depresión, cuando acompaña a otras enfermedades (diabetes, infección por el VIH, etc.)[22].

Factores relacionados con el paciente

La educación sobre la enfermedad y la información pertinente y clara sobre las razones para el tratamiento son vitales para generar adherencia. Si el paciente no conoce su propia enfermedad, o si no sabe las consecuencias, o si no está suficientemente motivado, no tendrá una buena aceptación de cuanto se le recomiende, y será indisciplinado con las prescripciones.

Se han propuesto múltiples intervenciones, como sesiones de educación, instrucciones escritas, reforzamientos por teléfono o por correo electrónico, así como educación e información a los acompañantes, y sesiones en grupo de pacientes con enfermedades similares[23].

Algunas características del paciente aumentan el riesgo de fallar: por ejemplo, las limitaciones físicas o cognoscitivas o las alteraciones en la memoria, aspectos que siempre deben ser tenidos en cuenta a la hora de hacer una prescripción.

En el proceso, de acuerdo con el paciente, deben considerarse siempre las creencias religiosas o los aspectos culturales que puedan interferir con las intervenciones.

Factores relacionados con el médico y con el equipo de salud

La educación y la información que se le entreguen al paciente dependen, sobre todo, de la información que, a su vez, tenga el médico, y de su capacidad para transmitirla de manera sencilla, práctica y, al mismo tiempo, convincente.

En la educación médica, se debe reforzar la adquisición de habilidades de comunicación, pues al dar instrucciones incompletas, inapropiadas, muy complejas o excesivamente científicas, se crea una barrera para aceptar recomendaciones[24].

Debe explicarse también cuáles pueden ser las consecuencias de la toma parcial de los medicamentos, como el mal control y el riesgo de complicaciones en pacientes hipertensos; o la aparición de resistencias bacterianas, en el caso de las infecciones, o el costo aumentado, con beneficios reducidos, en el caso de las enfermedades crónicas que necesitan control estricto, como la diabetes.

Factores relacionados con la intervención

Múltiples estudios han mostrado que la adherencia y el cumplimiento disminuyen si hay efectos adversos de los medicamentos, o si el paciente no ha sido prevenido sobre su aparición, o si la medicación es incómoda de tomar o si tiene múltiples intervenciones simultáneas[25]. El sabor y la textura de las medicaciones, así como los horarios y las condiciones para la toma, son importantes. El exceso de condiciones (como medicarse antes de ingerir los alimentos, o media hora antes de otro medicamento, o esperar al menos dos horas para consumir lácteos, entre muchas otras) interfieren, y solo deben recomendarse cuando su efectividad esté claramente comprobada.

Se han estudiado estrategias para reducir las barreras relacionadas con la complejidad de las formulaciones, y se ha encontrado que el riesgo de incumplimiento disminuye de manera significativa al combinar medicamentos en una sola tableta, por ejemplo, en vez de administrar dos tabletas. Un metaanálisis reciente[26] halló que en sujetos hipertensos, el uso de una combinación de dosis fija mejoró significativamente las posibilidades de adherencia (1,21 [IC95 %: 1,03-1,43]; $p = 0,02$), aunque no hubo una mejoría significativa en las cifras de presión arterial.

Factores relacionados con el sistema de salud

Aunque son los primeros mencionados por muchos médicos cuando se les interroga sobre causantes de no adherencia, los estudios han mostrado que no son la causa más frecuente[27-29], aunque sí son una causa importante, y de mucho mayor peso en las economías en desarrollo.

Entre los obstáculos para el cumplimiento se encuentran las demoras en atención, la entrega de medicamentos fuera del tiempo establecido, las entregas parciales, el costo de los medicamentos frente al poder adquisitivo y las dificultades para acceder a consultas y a oportunidades de renovación de fórmulas.

ESTRATEGIAS DE MEJORÍA

Se debe facilitar a toda costa que las indicaciones recomendadas se cumplan, y que los pacientes participen de manera activa en el proceso de toma de decisiones, aun cuando se trata de experimentos clínicos, en los que la participación debe ser limitada por razones del protocolo y del estudio en sí. Pero debe buscarse que los pacientes conozcan tanto como se pueda sobre su condición y sobre la intervención. Las estrategias propuestas deben estar siempre ajustadas a las particularidades de cada caso, tanto del paciente en sí como de su condición o su enfermedad y de su entorno[3].

Debe recordarse que hay una clara relación entre la adherencia y el efecto de los medicamentos, o *impacto positivo en el control de las enfermedades y en los desenlaces*[30], y que debe ponerse igual cuidado en la formulación de recomendaciones y en las medidas que faciliten y garanticen la concordancia y la persistencia.

La educación juega un papel de gran importancia: tanto sobre la naturaleza y las consecuencias de la enfermedad como sobre la medicación, sobre su potencial impacto y sobre las consecuencias de no apegarse a las recomendaciones, al igual que sobre los posibles eventos adversos y las formas posibles de controlarlos.

El paciente debe ser instruido sobre los signos y los síntomas de alerta para identificar de manera temprana las complicaciones y los efectos adversos, así como sobre las intervenciones para disminuir los riesgos de efectos adversos o de interacciones.

En cuanto a la dosificación y las recomendaciones para tomar el medicamento, se ha recomendado la búsqueda de regímenes tan simples como sea posible, con la menor cantidad de tabletas y las posologías más sencillas que se puedan. Es importante ayudar al paciente con

recomendaciones de medicamentos accesibles, haciendo especial énfasis en la disponibilidad y el costo[28].

Decidir sobre la formulación debe ser siempre el resultado de un acuerdo entre el profesional de la salud y el paciente (idealmente, también su familia cercana) para que el régimen sea tan específico como sea posible, y se ajuste de la mejor manera al estilo de vida y las posibilidades del sujeto.

En algunos casos debe enseñarse a la persona a ser no adherente, pero con inteligencia. Si se le enseñan los efectos negativos que puede tener un medicamento (hipoglucemia con los antidiabéticos, o sangrado con los anticoagulantes), debe también enseñársele cuándo suspender los medicamentos, si hacerlo de manera temporal o definitiva, y cuándo consultar. La confianza que tenga el paciente en su medicamento y en los posibles riesgos se relaciona directamente con su capacidad de cumplimiento.

Debe recomendarse, también, el uso de pastilleros con días de la semana y con la posibilidad de emitir alarmas o mensajes fijos o pregrabados[31]. En la revisión sistemática, publicada en 2011, acerca de la efectividad de utilizar paquetes con recordatorios, se incluyeron 12 estudios con 2196 participantes, y se encontró que dicha medida mejoraba, en promedio, el 11 % (IC 95 % 6 %-17 %) el porcentaje de tabletas ingeridas. Aunque los estudios eran muy heterogéneos, se pudo constatar que con el uso de esos paquetes con recordatorios se lograron mejores cifras tensionales diastólicas y mejores valores de hemoglobina glucosilada. En uno de los estudios se encontró que dicha ayuda fue la preferida por pacientes con bajo nivel de alfabetismo.

Tanto el personal de apoyo como el personal de salud diferente del personal médico deben estar entrenados no solo sobre los beneficios y los riesgos de las intervenciones, sino sobre las estrategias de comunicación y de educación con las que puedan mejorar las probabilidades de cumplimiento.

Los sistemas de salud deben buscar mejorar la oportunidad de consulta, mejorar las condiciones de entrega de los medicamentos, promover la continuidad en el tratamiento y buscar que haya la menor variabilidad posible en el personal que atiende a un paciente, así como promover las intervenciones psicosociales para estimular el cumplimiento.

Hay algunos recursos de utilidad potencial, que deben tenerse en cuenta por parte de las instituciones de salud, tales como la formación de organizaciones comunitarias, de grupos de pacientes y de programas de educación y difusión, así como de apoyo social. Los grupos de enfermedades específicas (asociaciones de pacientes con Ca, con diabetes, etc.) y de grupos de medicación (asociaciones de pacientes en quimioterapia, por ejemplo) pueden ser también muy útiles en enfermedades crónicas. También, los grupos de apoyo a la lactancia materna y los de alcohólicos anónimos.

Además de los programas de educación y de apoyo, debe promoverse el uso de recordatorios mediante llamadas telefónicas, alarmas electrónicas, el apoyo de familiares y vecinos y el uso de dispositivos electrónicos de recuerdo[32]. Debe tenerse en mente, sin embargo, que no siempre las intervenciones recomendadas tienen efectos significativos. Una revisión de Cochrane[33] encontró que solo la mitad de las intervenciones probadas en experimentos clínicos mejoraron la adherencia, y menos de un tercio mejoraron los desenlaces, si bien las intervenciones que se probaron eran complejas y constaban de varias intervenciones.

En algunos casos (particularmente, en la participación en estudios de investigación), es posible ofrecer incentivos económicos o recompensas por la participación y el cumplimiento. No hay ningún obstáculo ético para su empleo, siempre y cuando lo entregado tenga proporcionalidad con los esfuerzos, las dificultades o las incomodidades a las que se vea sometido el participante.

En casos como la tuberculosis, en la que es tan importante mantener el nivel óptimo de tratamiento, para evitar la aparición de resistencias bacterianas, se han propuesto estrategias como la terapia con observación directa (DOT, por las siglas en inglés de *Directly Observed Therapy*) o la terapia acortada estrictamente supervisada (TAES). Y también en esos casos, así como en otras enfermedades crónicas asociadas a bajos recursos económicos, se propone la entrega de incentivos económicos, o, mejor aún, suministrar alimentos básicos y suplementos nutricionales.

Como reglas generales, las intervenciones evaluadas no produjeron beneficios que se sostuvieran por más de seis meses (lo que habla de la necesidad de reforzarlas de manera periódica, y, eventualmente, tener que combinar o alternar

estrategias). Pero es claro que la simplificación de los regímenes (una vez al día *vs.* varias, por ejemplo) mejora la adherencia.

Se estima que el costo de la no adherencia a las medicaciones le cuesta a Estados Unidos 100 billones de dólares por año, y que la no adherencia a las estatinas, a la aspirina y a los betabloqueadores en los pacientes con infarto reciente triplica las tasas de mortalidad luego de la salida del hospital. Estos datos, similares a los obtenidos en otras condiciones, refuerzan la necesidad de buscar activamente la mejoría en la adherencia durante todo el periodo de la enfermedad o la condición.

MEDICIÓN DE LA ADHERENCIA

Es muy importante saber cuál es el grado de cumplimiento del paciente con las recomendaciones, y cuánto del medicamento toman, en general, para poder decidir si lo que puede necesitarse en caso de no lograrse el control apropiado de la condición sea modificar la dosis, adicionar otro medicamento o un cambio total.

Las mediciones pueden hacerse mediante diarios que se le pide al paciente que lleve, o mediante información de la familia, el conteo de tabletas en las citas de control, los dispositivos electrónicos que registran la apertura del frasco del medicamento, o la medición de niveles sanguíneos o en la orina, entre otros.

Hay experiencias exitosas que pueden tomarse como ejemplo, tales como las DOT y las TAES ya mencionadas para la tuberculosis, o las existentes para los pacientes con infección por el VIH.

CONCLUSIONES

No hay una fórmula única para conseguir la adherencia en estudios de investigación o cuando se trata a pacientes. Es fundamental evaluar las estrategias utilizadas y buscar información sólida sobre la efectividad demostrada de las técnicas potenciales para mejorar la adherencia de lo que frecuentemente hacemos.

Debe buscarse siempre identificar las barreras a la adherencia y a la finalización de los tratamiento, y recordar que una de las más poderosas es la buena comunicación entre los investigadores o los proveedores en salud y los pacientes, para lograrlo.

En la conducción de un estudio o en el tratamiento de un paciente, es fundamental tener en cuenta que los pilares para lograr la adherencia son la educación, el refuerzo y la permanente reevaluación de las estrategias.

Referencias

1. Granger BB, Swedberg K, Ekman I, for the CHARM investigators. Adherence to candesartan and placebo and outcomes in chronic heart failure in the CHARM programme: double-blind, randomised, controlled clinical trial. Lancet. 2005;366:2005-11.
2. Horwitz RI, Viscoli CM, Berkman L, et al. Treatment adherence and risk of death after a myocardial infarction. Lancet. 1990;336:542-5.
3. Irvine J, Baker B, Smith J, et al. Poor adherence to placebo or amiodarone therapy predicts mortality: results from the CAMIAT study. Psychosom Med. 1999;61:566-75.
4. Anonymous. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med. 1980;303:1038-41.
5. White HD. Adherence and outcomes: it's more than taking the pills. Lancet. 2005;366:1989-91.
6. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med. 2005;353:487-97.
7. Knoops KT, de Groot LC, Kromhout D, et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA. 2004;292: 1433-9.
8. Chapman RH, Benner JS, Petrilla AA, et al. Predictors of adherence with antihypertensive and lipid-lowering therapy. Arch Intern Med. 2005;165:1147-52.
9. Simpson SH, Eurich DT, Majumdar SR, et al. A meta-analysis of the association between adherence to drug therapy and mortality. BMJ. 2006;333:15-20.
10. McDonald K, Ledwidge M, Cahill J, et al. Elimination of early rehospitalization in a randomized, controlled trial of multidisciplinary care in a high-risk, elderly heart failure population. Eur J Heart Fail. 2001;3:209-15.
11. van der Wal MH, Jaarsma T, Van Veldhuisen DJ. Non-compliance in patients with heart failure; how can we manage it? Eur J Heart Fail. 2005;7:5-17.

12. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care. 1975;13:10-24.
13. Aronson JK. Compliance, concordance, adherence. Br J Clin Pharmacol. 2007;63:383-4.
14. Bosworth HB. Medication adherence. En: Bosworth HB, editor. Improving patient treatment adherence: A Clinician's Guide. New York: Springer; 2010.
15. Haynes RB, Ackloo E, Sahota N, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev. 2008:CD000011.
16. Miller NH, Hill M, Kottke T, et al. The multilevel compliance challenge: recommendations for a call to action: a statement for healthcare professionals. Circulation. 1997;95:1085-90.
17. Salimbene S. What language does your patient hurt in? A practical guide to culturally competent patient care. Amherst, MA: Diversity Resources; 2005.
18. Spector R. Cultural diversity in health and illness, 5h ed. Upper Saddle River, NJ: Prentice Hall; 2000.
19. DiMatteo MR, Haskard KB, Williams SL. Health beliefs, disease severity, and patient adherence: a meta-analysis. Med Care. 2007;45:521-8.
20. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care. 2004;42:200-9.
21. DiMatteo MR, Giordani PJ, Lepper HS, et al. Patient adherence and medical treatment outcomes: a meta-analysis. Med Care. 2002;40:794-811 .
22. González JS, Peyrot M, McCarl LA, et al. Depression and diabetes treatment A meta-analysis. Diabetes Care. 2008;31:2398-403.
23. Segadora J, Gil-Guillen VF, Orozco D, et al. The effect of written information on adherence to antibiotic treatment in acute sore throat. Int J Antimicrob Agents. 2005:26:56-61.
24. Brand PL, Klok T, Kaptein AA. Using communication skills to improve adherence in children with chronic disease: the adherence equation. Paediatr Respir Rev. 2013;14:219-23.
25. Ingersoll KS, Cohen J. The impact of medication regimen factors on adherence to chronic treatment: a review of literatura. J Behav Med. 2008;31:213-24.
26. Gupta AK, Arshad S, Poulter NR. Compliance, safety, and effectiveness of fixed-dose combinations of antihypertensive agents: a meta-analysis. Hypertension. 2010;55:399-407.
27. McKenzie SJ, McLaughlin D, Clark J, et al. The burden of non-adherence to cardiovascular medications among the aging population in Australia: a meta-analysis. Drugs Aging. 2015;32:217-25.
28. Maxwell CJ, Stock K, Seitz D, et al. Persistence and adherence with dementia pharmacotherapy: relevance of patient, provider, and system factors. Can J Psychiatry. 2014;59:624-31.
29. Fiander M, McGowan J, Grad R, et al Interventions to increase the use of electronic health information by healthcare practitioners to improve clinical practice and patient outcomes. Cochrane Database Syst Rev. 2015:CD004749.
30. Ruppar TM, Dobbels F, Lewek P, et al. Systematic review of clinical practice guidelines for the improvement of medication adherence. Int J Behav Med. 2015 Mar 25. [Publicación electrónica anticipada, última consulta 28 de marzo de 2015].
31. Mahtani KR, Heneghan CJ, Glasziou PP, et al. Reminder packaging for improving adherence to self-administered long-term medications. Cochrane Database Syst Rev. 2011:CD005025.
32. Lieber BA, Taylor B, Appelboom G, et al. Meta-analysis of telemonitoring to improve HbA1c levels: Promise for stroke survivors. J Clin Neurosci. 2015:S0967-5868(14)00706-1. [Publicación electrónica anticipada, última consulta 28 de marzo de 2015].
33. Nieuwlaat R, Wilczynski N, Navarro T, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev. 2014:CD000011.



SECCIÓN 7

Recolección, análisis y presentación de datos

CAPÍTULO 28 Desarrollo de cuestionarios
Carlos Gómez-Restrepo, María Beatriz Ospina García

CAPÍTULO 29 Plan de análisis
María Nelcy Rodríguez Malagón, Juan Gabriel Ruiz Peláez

CAPÍTULO 30 Modelado
Jhon Camacho

CAPÍTULO 31 Construcción de bases de datos
Jhon Camacho

CAPÍTULO 32 Preparación de datos
Jhon Camacho

CAPÍTULO 33 Presentación estadística de datos
Martín Alonso Rondón Sepúlveda

CAPÍTULO 34 Medicina basada en evidencia
Álvaro J. Ruiz Morales

CAPÍTULO 35 Apreciación crítica de la literatura
Juan Manuel Lozano León

CAPÍTULO 36 Principios éticos en investigación clínica y epidemiológica
Álvaro J. Ruiz Morales, Fernando Suárez Obando

CAPÍTULO 37 Autoría
Álvaro J. Ruiz Morales

CAPÍTULO 38 Informática biomédica y epidemiología clínica
Fernando Suárez Obando



CAPÍTULO

28

Desarrollo de cuestionarios

Carlos Gómez-Restrepo • María Beatriz Ospina García

INTRODUCCIÓN

Desde hace algunos años, en las ciencias sociales y, más recientemente, en el campo de la salud, se ha destacado la importancia de la medición. La necesidad de recuperar datos y de medir variables complejas, tales como la calidad de vida, el dolor, la inteligencia, la personalidad, los estados de ánimo y el bienestar, entre otras, ha estimulado grandes adelantos en este campo. Tal desarrollo se produjo no solo de la mano de los avances en sociología, psicología, educación y antropología, sino también, de la estadística. En el campo médico, probablemente, fue la psiquiatría la pionera en la exploración de dichas áreas.

Habitualmente, se han clasificado las variables en salud, según el proceso de medición, como ***variables duras*** **y** ***variables blandas.***

Variables duras son las que pueden ser medidas con instrumentos precisos, en las que los resultados tienen poca posibilidad de influencia externa y son repetibles.

EJEMPLO

Determinar el colesterol sérico o determinar los niveles de ácido úrico son procesos objetivos, y al repetir las mediciones en condiciones similares se obtienen resultados iguales o muy similares, pues la evaluación se hace con máquinas precisas. Otros ejemplos de variables duras son: determinar la litemia, determinar el recuento mecánico de eritrocitos en la sangre, medir la longitud de una extremidad, tomar la presión arterial, etc.

El concepto de variable dura no implica que sean siempre válidas las mediciones, o que no pueda haber errores o que sean de mejor calidad que las blandas. Tiene que ver con el proceso de medición; una variable dura puede considerarse, en general, más simple de medir que una blanda, pero no tiene características inherentes de calidad por el hecho de ser dura.

Variables blandas son las que no son objetivas o que tienen un alto grado de subjetividad. Se trata de aspectos cuya medición es más compleja, pues en ocasiones no se tienen instrumentos de medición definidos, inequívocos o estandarizados para evaluarlas e interpretarlas.

EJEMPLO

La evaluación del dolor postoperatorio presenta un reto difícil para el anestesiólogo, por cuanto hay una gran variabilidad en el umbral al dolor, así como la influencia de factores externos e internos en la percepción de dolor o en su manifestación, y no hay instrumentos que permitan hacer tal evaluación. El dolor posquirúrgico puede atenuarse significativamente si el paciente recibe la noticia de que la cirugía mostró que no tenía cáncer. O el dolor posterior a una cesárea puede disminuir significativamente cuando se le entrega el bebé a la madre.

Las variables blandas no son variables de menor importancia, ni son inapropiadas ni poco correctas ***per se***. Los únicos conceptos inherentes están relacionados con la dificultad para medirlas o para estandarizar su proceso de evaluación y de repetibilidad (en ocasiones, poco predecible).

Los términos ***duro*** y ***blando*** llevan, incorrectamente, a algunas personas a clasificar lo que corresponde a duro como importante, y lo que corresponde a variables blandas, como secundario, menos científico o menos importante. Se trata de un error bastante común, que debe evitarse y corregirse.

En algunos casos, hay variables que no corresponden exactamente a la clasificación anterior, pues tienen apariencia dura, pero características blandas. Una de las variables utilizadas más a menudo como indicador de calidad de atención es la duración de la estancia hospitalaria, o días en cuidado intensivo o tiempo hasta el retorno al trabajo. Por tratarse de una medición cuyo resultado es numérico, aparenta ser objetiva y medible con métodos estandarizados. Sin embargo, la decisión de dar de alta a un paciente o de trasladarlo a cuidado intensivo tiene una enorme influencia de factores externos a la condición de salud del individuo (aspectos económicos, logísticos de la institución, sociales, etc.), y cuando se lee un estudio que evalúa un tratamiento nuevo mediante días de estancia, debe considerarse que si bien hay métodos estandarizados para hacer la medición, la variable tiene per se una enorme susceptibilidad a influencias externa, y, por ende, una posible variabilidad significativa en su determinación.

En términos generales, en las variables blandas se recurre a estrategias como la conversión de conceptos cualitativos (dolor) en aspectos cuantitativos (escala de dolor de uno a diez), al desarrollo de cuestionarios o a la elaboración de constructos. En el presente capítulo nos detendremos en la evaluación de estos últimos.

Cada vez más a menudo en ciencias de la salud, la preocupación de los equipos de salud tiene que ver con la cantidad en su relación con la calidad. Ya no es importante tan solo aumentar la expectativa de vida: también lo es aumentarla, pero acompañada de una buena calidad de vida. Reducir con medicamentos la presión arterial, del colesterol o de la glucemia es importante, siempre y cuando se logre sin alterar el sueño, el ánimo, la función sexual u otros aspectos vitales. Disminuir el número de recaídas en una enfermedad bipolar tiene menos sentido si se acompaña de alteraciones significativas en la capacidad para las relaciones interpersonales, en la independencia, en la capacidad de reacción o en el afecto. Medir estos aspectos, difícilmente cuantificables, ha generado mucho interés y ha propiciado el desarrollo de métodos que solucionen las dificultades naturales para evaluarlos.

Los investigadores suelen encontrarse con dificultades en la elaboración de un protocolo; particularmente, en la definición de variables que corresponden a la categoría de variables blandas, o en la redacción de cuestionarios o encuestas, así como en el desarrollo y la planeación de análisis para escalas.

Se busca con el presente capítulo revisar algunos de los aspectos que debe tomar en cuenta el investigador[1-4] cuando desarrolla un cuestionario.

Cuestionarios

Un cuestionario puede definirse como un documento escrito que se utiliza para solicitar a uno o más lectores o escuchas contestar una o más preguntas (puntos)[5].

Los cuestionarios también se han definido como instrumentos escritos de medición que buscan recolectar datos generales de los encuestados, así como evaluar sus opiniones, sus actitudes, sus creencias, sus comportamientos o sus atributos[6].

Las actitudes y las creencias son, por lo general, más sensibles a variaciones en la redacción

de las preguntas y a las palabras utilizadas, lo cual debe ser cuidadosamente considerado cuando se explore este tipo de dimensiones.

Los cuestionarios son fundamentales en toda investigación, pues son la base para recolectar los datos que se requieren para analizar un estudio.

Pueden cubrir desde aspectos más bien generales, como describir las características de los encuestados (sexo, edad, condición socioeconómica, estado civil, escolaridad, etc.) hasta datos más específicos, como los niveles de dolor, la evaluación del estado de salud, y la apreciación sobre la calidad de la atención brindada, entre otros.

En tal sentido, podemos decir que un buen estudio debe tener instrumentos estructurados para obtener los datos, y dentro de las formas de recolección de datos, los cuestionarios son de vital importancia, pues incrementan la validez de los resultados.

Un estudio perfectamente elaborado en su diseño puede resultar por completo inválido si los cuestionarios no captan lo que en realidad expresa el participante[7,8].

De acuerdo con su forma de administración, los cuestionarios pueden ser ***autoadministrados*** o ***administrados por terceros***.

Los cuestionarios autoadministrados tienen la ventaja de que pueden ser entregados por un tercero sin calificaciones o sin entrenamiento especial; además, pueden ser distribuidos masivamente por correo o por medios electrónicos. Como desventajas, tienen que: se pierde la oportunidad de asegurar que se contesta todo el cuestionario; se pierden aspectos importantes de la reacción durante las respuestas; no hay oportunidad de interacción directa con el investigador, quien puede proporcionar información útil, y se necesita un nivel de educación suficiente para la lectura, la comprensión, el análisis y la escritura o la consignación de las respuestas.

Los cuestionarios administrados por terceros necesitan entrenamiento formal para el encuestador, por cuanto no solo se trata de obtener las respuestas, sino de no influir en ellas con el lenguaje ni con la forma de enunciar las preguntas; se debe ser en extremo cuidadoso en el comportamiento durante la entrevista, y evitar comentarios, reacciones a las respuestas o proporcionar ayudas. Como ventajas, pueden ser administrados a personas con bajo un nivel de lectura o analfabetas, así como a personas con déficit sensorial o de atención; adicionalmente, se puede controlar el tiempo de la entrevista, asegurarse de que se leen y se contestan todas las preguntas y evaluar inmediatamente que no haya errores o faltantes.

Al momento de diseñar un cuestionario debe pensarse en varios puntos:

Antes de iniciar el estudio, y según la pregunta y los objetivos del estudio, el investigador debe precisar la información que requiere para su desarrollo, con el fin de pensar y planear los cuestionarios o los instrumentos que se requieran.

A la vez, deberá pensar en preguntas para el cuestionario que resuelvan las dudas planteadas, que respondan a los objetivos del estudio, que sean específicas, no redundantes, y útiles para la investigación.

En general, se tienen cuestionarios con preguntas ***abiertas*** o ***cerradas***, o de tipo mixto. Las preguntas abiertas son más difíciles de agrupar, tabular, precisar, contestar y analizar. Tienen la ventaja de dejar expresarse más al participante y motivarlo más, pero también, la desventaja de que algunas personas tienen más dificultades para responder las preguntas; además, suelen requerir un mayor tiempo de administración, y, sobre todo, dificultan la tabulación y el análisis de los datos, lo que puede acarrear mayores tiempos y mayor consumo de recursos.

Las preguntas cerradas son más precisas, tienen menores dificultades para contestarse y son más fáciles de agrupar, clasificar, tabular y analizar. No obstante, si no están bien formuladas, pueden no ofrecerle al encuestado todas las opciones de respuesta, y, por tanto, introducir sesgos de medición.

Los cuestionarios mixtos tienen características propias de ambos tipos de preguntas, y la ventaja potencial de tener mayor variedad y así garantizar la atención de los participantes. Esta sección se ampliará en el aparte "Cómo escribir buenas preguntas".

Redactar o recuperar las preguntas

Muchas de las preguntas que recogen información general del encuestado pueden obtenerse de instrumentos que han sido utilizados para el mismo propósito en otros estudios. Por ejemplo, para indagar sobre el nivel de escolaridad, el estrato socioeconómico o la composición familiar, se encuentran fácilmente formatos ya utilizados y probados. Para redactar las preguntas, hay distintos formatos, que serán comentados en detalle en el aparte "Cómo escribir buenas preguntas".

Ordenar las preguntas del cuestionario de una manera relevante en cuanto al conocimiento

Una vez se tengan las preguntas, debe procurarse organizarlas por secciones, según temáticas, lo que las hace más fáciles de seguir, comprender y responder. Por lo general, se parte de lo general a lo particular, de lo fundamental, o requerido, a lo secundario, o derivado. Por ejemplo, es indispensable preguntar si el padre está vivo antes de indagar acerca de la relación del encuestado con este. **Por otra parte,** se recomienda **poner primero las preguntas de** mayor relevancia para mantener la atención del participante. Si el cuestionario es largo, ello permite obtener respuestas fundamentales con mayor colaboración y concentración antes de que haya cansancio o desatención por parte del participante. Sin embargo, en casos donde haya preguntas sensibles (por ejemplo, sobre maltrato, abuso, actividad sexual, disputas familiares, opiniones sobre el jefe, etc.), estas se pueden situar un poco después, para evitar que la persona se resista y, de alguna manera, tenga menor confianza al responder. En ocasiones, y con el fin de evaluar la veracidad de las respuestas, se repiten preguntas de una sección en otra, para así obtener información sobre la validez de las respuestas.

Escribir la introducción al cuestionario

Una vez las preguntas han sido elaboradas, debe escribirse una sección de introducción en el cuestionario, donde se le explique de manera breve al participante cuáles son el objetivo y el propósito (para qué) del cuestionario, la necesidad y la importancia de su colaboración, así como el tiempo estimado para su aplicación.

Probar el cuestionario



Una vez el cuestionario ha sido elaborado, es indispensable diagramarlo, y que por lo menos dos pares provean retroalimentación acerca de este, para mejorar cuanto se requiera. Conviene que las personas encargadas de evaluar el cuestionario sean ajenas a su construcción, que lo lean y hagan críticas constructivas sobre los cambios que sean necesarios.

También se le puede leer el cuestionario a un grupo de personas (tres a cuatro), semejantes a quienes van a contestarlo, y preguntarles qué entienden en cada punto y qué dificultades vieron para comprender los términos y las preguntas. Esto da una idea sobre la comprensión global, sobre los tiempos de aplicación u otros comentarios que surjan de este grupo, y que sirvan para mejorar el cuestionario. Tal estrategia puede ayudar sobremanera a mejorar el instrumento.

Obtener retroalimentación del cuestionario elaborado

Es importante tener a disposición un lugar cómodo, con poco ruido y buena luz para aplicar el cuestionario al grupo de personas (dos a cuatro) que darán retroalimentación al instrumento elaborado. Una vez se tengan los comentarios, estos se incorporan a la nueva versión, en caso de que así convenga. Siempre debe hacerse un piloto con personas semejantes a quienes participarán en el estudio; dichos pilotos son particularmente necesarios y útiles en estudios de gran envergadura, como encuestas regionales o nacionales.

Planeación de la aplicación del cuestionario

En general, esta parte se incorpora en la conducción del estudio. Se deben planear los requerimientos para aplicar el cuestionario durante el estudio, deben escribirse los manuales de aplicación para los encuestadores y las instrucciones que describan los pasos, pregunta por pregunta, en el caso de cuestionarios complejos o largos. Así mismo, se deben contemplar

detalles logísticos como qué se requiere para contestar el cuestionario (bolígrafo, colores, etc.), cómo enviar los cuestionarios, cómo numerarlos, quién y cómo los entregará, el tiempo de la encuesta, dónde se archivarán, y otros detalles que lleven a una recolección idónea de la información.

Cómo escribir buenas preguntas

Algunos autores, como Woodward ***et al.*** y Streiner, proponen reglas para la formulación de preguntas y describen algunos de los errores más comunes. A continuación, algunas de sus recomendaciones:

- Las palabras empleadas deben ser simples, directas y familiares. Se debe evitar el uso de abreviaturas, frases foráneas (extranjerismos) y terminología técnica. Así mismo, se deben adaptar las preguntas a la población a la que se dirijan.

EJEMPLO

Dirigida al público, podría plantearse esta pregunta: ¿Ha sentido dolor de espalda? Sin embargo, para un médico puede formularse como: ¿Ha presentado lumbago?, que puede ser entendida por un trabajador de la salud, pero tal vez no por otras personas.

También en el campo médico, ***¿Tiene usted hipertensión?*** **puede ser entendida como tener** la mencionada enfermedad o como tener muy alta tensión en cierto momento (estar muy tensionado, estresado u otra). ***¿Ha estado deprimido?*** **puede ser interpretado como angus**tiado, miedoso, con problemas, etc.
¿Ha sido diagnosticado con ***EPOC?***: **deben** evitarse las siglas y escribir o decir: ***enfermedad pulmonar obstructiva crónica.***

- Las preguntas deben ser claras y específicas. Algunas preguntas como **"¿Consume usted** alcohol habitualmente?" **no son claras, por**que quienes responden pueden interpretar "habitualmente" de muchas maneras. Para algunos puede ser 20 veces al mes, para otros 3 veces al día, o 2 veces por semana. Por tal motivo, es fundamental precisar este concepto dentro de la pregunta, o escribirla de manera diferente: por ejemplo, indicando el número de días que se ha consumido alcohol durante los últimos 15 días.
- Debe evitarse la formulación de preguntas dobles.

Con el ánimo de ahorrar tiempo y espacio, a veces se formulan dos preguntas al mismo tiempo. Eso puede llevar a dificultades o equívocos en la interpretación de la respuesta.

EJEMPLO

¿Está usted de acuerdo o en desacuerdo con la atención que le brindan el sistema de seguridad social y el hospital a los que está afiliado?

De acuerdo _____ En desacuerdo ____

En este caso, podría ser que el entrevistado esté de acuerdo con el sistema de seguridad social y en desacuerdo con el hospital, o viceversa; pero dado el tipo de pregunta, ello no puede ser esclarecido. Errores de este tipo son comunes en medicina:

EJEMPLO

¿Ha estado triste y lloroso? Sí ____ No ____
¿Tiene eritema y dolor? Sí ____ No ____
¿Tiene los ojos rojos y le arden? Sí ____ No ____
¿La atención es buena y oportuna? Sí ____ No ____

En todas las anteriores preguntas vemos que puede haber más de una opción de respuesta. Debe recordarse que se deben prever todas las posibilidades de respuestas, lo cual incluye las opciones de "todas las anteriores", "ninguna de las anteriores", "no aplica", "no sabe/no responde", "no hay datos" y "no se sabe por qué", etc.

- No deben escribirse preguntas con dobles negativos, que sean confusas de leer y de entender.

EJEMPLO

¿La falta de atención psiquiátrica no es un problema en la zona del Pacífico colombiano?

Sí ____ No____

Podría mejorarse así:

¿La falta de atención psiquiátrica es un problema en la zona del Pacífico colombiano?

Sí ____ No____

- Debe evitarse formular preguntas sesgadas, tendenciosas o con palabras cargadas afectivamente.

EJEMPLO

¿Cómo le parece la mala atención de este hospital?

En este caso, podría evitarse la palabra "mala".

EJEMPLO

¿A menudo consulta al médico por problemas triviales?
¿Los salarios son muy bajos?

- No debe sesgarse la pregunta al dar solo algunas alternativas de respuesta.

EJEMPLO

¿Usted, como médico, trabaja en un hospital o en la práctica privada?

En este caso, no se dan todas las opciones laborales de los médicos (empresas, centro de salud, administración, etc.), y se mencionan dos que pueden presionar para que se escoja entre ellas.

EJEMPLO

¿El director del hospital es muy amable o amable?
¿No es verdad que se siente mejor?

Estas formas de elaborar las preguntas pueden sesgar las respuestas y forzar al participante a contestar que, por lo menos, "es amable", en la primera pregunta; en la segunda habrá gran presión para contestar que sí, que se siente mejor.

- Debe evitarse la formulación de preguntas que afecten la autoestima o sean vergonzosas para el participante.

EJEMPLO

Para una adolescente que quisiera tener novio y no lo ha tenido, si se le pregunta: *¿ha tenido novio?*

Esta pregunta podría no ser contestada con sinceridad.
Podría contestarse idóneamente si se le pregunta:

EJEMPLO

¿Usted tiene novio actualmente? o ¿Durante los últimos seis meses ha tenido usted novio?

- En lo posible, evite preguntas personales o que desencadenen implicaciones negativas.

EJEMPLO

¿Cuánto es su ingreso mensual? _____ pesos.

En caso de requerirse este tipo de preguntas, procure poner rangos:

EJEMPLO

¿Cuál de las siguientes categorías describe mejor su salario mensual?

a. Menos de $250 000.
b. Entre $251 000 y $500 000.
c. Entre $500 001 y $1 000 000.
d. Entre $1 000 001 y $2 000 000.
e. Más de $2 000 000.

- No haga preguntas que, probablemente, desencadenarán respuestas poco o nada sinceras.

EJEMPLO

¿A menudo pregunta en clase cosas triviales?
Sí ____ No____

La mayoría de las personas responderán que no.
Es posible que esta pregunta, hecha de otra forma, propicie respuestas más honestas:

EJEMPLO

¿Las preguntas triviales aumentan el conocimiento?

- Debe clarificarse el tiempo requerido en la respuesta.

EJEMPLO

¿Durante el último año ha sido hospitalizado?
Sí ____ No ____

¿Qué significa "el último año"? Quizá haya al menos tres interpretaciones para ello: puede significar "en los últimos más o menos doce meses", o "desde el primero de enero hasta ahora" o, si ya se está en junio, "desde junio del año pasado".

EJEMPLO

Podría mejorarse así:
¿Durante los últimos 12 meses ha sido hospitalizado?
Sí ____ No ____

- No deben utilizarse palabras que tengan dos o más significados.

EJEMPLO

¿Ha visto al médico en los últimos 30 días?
Sí ____ No____

La palabra "visto" puede significar visitarlo para consulta o verlo en algún sitio diferente de la consulta.

- Deben evitarse las preguntas con negaciones.

EJEMPLO

¿Nunca se siente bien en el hospital?

Puede obtenerse mayor validez en las respuestas si no se usan palabras como: ***rara vez,*** *no* **y** ***nunca.***

- Se debe prestar atención a la longitud de las preguntas.

Algunos autores han demostrado que cuanto más cortas sean las preguntas, mayor será la validez de la información que se obtenga. Las preguntas de entre 70 y 80 caracteres tienden a arrojar coeficientes de validez de 0,1 o menores, en tanto que las que tienen entre 10 y 20 letras tienden a tener coeficientes 4 veces mayores[9]. Por tal motivo, se recomienda construir preguntas cortas, sin que ello comprometa su claridad.

- *Debe* **controlarse cuidadosamente el nivel de** lectura del entrevistado.

Como regla general, se cree que una pregunta debe ser comprendida por un niño de 12 años. En los países en vías de desarrollo se debería bajar esa cifra a 10 años (alrededor del grado quinto de primaria), con el fin de obtener una mayor comprensión en poblaciones con bajo nivel de escolaridad. A la vez, se debe procurar el uso de palabras que no generen ambigüedad ni sean problemáticas o tengan múltiples significados.

Tipos de preguntas

De acuerdo con la formulación de las preguntas, estas pueden ser abiertas o cerradas. Cada una de dichas opciones tiene sus indicaciones y sus contraindicaciones, así como sus ventajas y sus desventajas.

Preguntas abiertas

Permiten que la persona elabore con mayor libertad y espontaneidad sus respuestas. En términos generales, pueden tener las siguientes ventajas y desventajas:

Ventajas

- Estimulan el pensamiento.
- Sirven para explorar un tema y obtener sugerencias.
- Son útiles para preparar preguntas cerradas.

Desventajas

- Requieren que la persona sepa escribir (en caso de que no le sean leídas por terceros).
- Para algunas personas, puede resultar difícil crear, articular o redactar lo que piensan.
- Dependen más de la memoria de la persona.
- El análisis estadístico de los datos obtenidos con ellas es difícil.
- Las respuestas pueden ser difíciles de agrupar, de tabular y de analizar.

- Algunas respuestas pueden ser difíciles de leer o ininteligibles.
- Puede haber respuestas incompletas o irrelevantes.
- Su tabulación y su análisis son más costosos y dispendiosos.

Preguntas cerradas

En este tipo de preguntas, se ofrecen varias opciones de respuestas seleccionadas previamente por el investigador. Las opciones pueden ser tan sencillas como escribir "sí" o "no", "de acuerdo" o "en desacuerdo", u otras, como las escalas visuales o de adjetivos, que revisaremos a continuación.

Ventajas

- Las preguntas son más fáciles de contestar, y mediante respuestas precisas.
- Las respuestas se pueden agrupar, clasificar, tabular y analizar con mayor facilidad.
- Las respuestas son susceptibles de análisis estadísticos y de representación gráfica.
- Se pueden determinar el grado de compromiso, la intensidad de respuesta y la frecuencia.
- Sus respuestas son más fáciles de interpretar que las respuestas abiertas.
- La elaboración de sus bases de datos tiene menores dificultades.

Desventajas

- Se limita la posibilidad de obtener respuestas diferentes de las formuladas.
- La pregunta que tiene respuestas prefijadas debe ser más cerrada, limitada y precisa, lo cual puede hacer que se pierdan matices o aspectos diferentes.
- Pueden limitarse o dejar de percibirse aspectos importantes de la problemática en estudio.
- Dentro de las preguntas cerradas existen dos tipos de respuesta:
- Categóricas.
- Continuas.

Respuestas categóricas

Como su nombre lo indica, utilizan categorías o juicios categóricos como respuesta.

EJEMPLO

¿Ha tenido dolor de cabeza hoy?

Sí ____ No____

Los médicos disminuyen la gravedad de la enfermedad.

Verdadero ____ Falso ____

¿Cuál de los siguientes síntomas tiene usted ahora?

Tristeza ____

Llanto ____

Dolor de cabeza ____

Este tipo de respuesta puede no ser apropiado para algunas variables como actitudes y comportamientos, que, por lo general, caen dentro de un continuo.

EJEMPLO

Ejemplo de ello sería preguntar:

¿Tiene usted dificultad para cocinar? Sí ____ No___

¿Está de acuerdo con el director? Sí ____ No ____

Estas preguntas podrían tener diferentes niveles de respuesta que no cabrían dentro de las opciones de respuesta categóricas sugeridas anteriormente, y se puede perder información relevante. Por otra parte, pueden crear cierta confusión para la persona que responde la pregunta y suscitar inexactitudes.

Respuestas continuas

En general, se ha determinado que hay una relación entre el número de opciones que se ofrece como respuesta (categorías, por ejemplo) y la confiabilidad de las respuestas.



Se ha determinado que las escalas que tienen entre cinco y siete opciones son las que ofrecen la mejor validez. Un menor número de categorías puede reducir la confiabilidad; a mayor número, se introduce más dificultad y se reduce la calidad de las respuestas.

Por otra parte, al tener un número impar de categorías se facilita que quien responde pueda ponerse en posición neutral, en el caso de escalas bipolares. En contraposición, tener un número

par de categorías obliga a los encuestados a ubicar sus respuestas en uno u otro extremo (*lado*). Para decidir cuál de ellas utilizar, conviene revisar los objetivos del estudio: si es importante saber hacia qué lado (*polo*) se ubican las respuestas de las personas, sería más adecuado contar con un número par de categorías. En caso contrario, podría optarse por un número impar de alternativas de respuesta. No obstante tener solo tres categorías lleva a que la mayoría de respuestas se den en la categoría de la mitad.

Dentro de los tipos de escalas de respuesta más comúnmente utilizadas se encuentran las siguientes:

Escala visual análoga (EVA)

Dada su simplicidad, es una de las más utilizadas en el campo de la salud. Consiste en una línea que, habitualmente, mide 100 mm, y en cuyos extremos se ponen palabras como: "no dolor" y en el otro "el dolor más fuerte imaginable". La persona que responde debe poner una X o una barra (|) en el lugar donde piense que se encuentra su respuesta. La evaluación de la respuesta se hace, habitualmente, midiendo en milímetros la ubicación de la respuesta.

EJEMPLO

¿Qué tan fuerte ha sido su dolor de cabeza hoy?

No dolor _____ El dolor más fuerte imaginable

Este tipo de respuestas pueden ser difíciles de entender en ciertos contextos, y hasta el 7 % de las personas no puede responderlas[10]. Las personas mayores, por ejemplo, tienden a entender mejor las escalas tipo termómetro y en posición vertical; en las poblaciones con niveles bajos de conocimiento, es difícil comprender el proceso.

Además de su simplicidad, las escalas **EVA** pueden utilizarse con facilidad para hacer mediciones de seguimiento. Si el sujeto marcó inicialmente su dolor, su tristeza o una disnea en una **EVA de 76 mm, es fácil hacer la comparación** con un nuevo valor (por ejemplo, 43 mm) luego de una intervención. A menudo, se estila anclar la respuesta de control preguntando al sujeto:

En la medición inicial usted marcó su dolor en este punto. ¿Cómo cree que ha cambiado?

Escalas con adjetivos

Se parece a la EVA, pero utiliza adjetivos dentro de la escala de respuesta (continua o discreta). Tiene la ventaja de ofrecer significados más claros e inequívocos dentro de la escala de respuesta.

EJEMPLO

En relación con las personas que conoce, ¿Cómo considera su rendimiento laboral, comparado con el promedio de las personas con las que trabaja?

Respuesta discreta

Muy por debajo del promedio	Debajo del promedio	Promedio	Encima del promedio	Muy por encima del promedio

Respuesta continua

I_____X_____I

Escala tipo termómetro

Similar a la EVA, pero con la escala ubicada verticalmente y con puntajes en los diferentes estadios.

Escala Likert

La única diferencia en relación con las anteriores es que utiliza opciones de respuesta que van de un extremo a otro (por ejemplo, desde estar totalmente de acuerdo hasta estar totalmente en desacuerdo). La escala de Likert (se pronuncia como está escrito; se la llama así por el apellido del psicólogo que la inventó: Rensis Likert) tiene la ventaja de que permite identificar tendencias y explorar distintos niveles de respuesta, lo cual puede resultar particularmente útil para temas o asuntos delicados o desafiantes. Si el número de opciones es impar permite (si ello es de interés para el estudio) identificar si hay una tendencia a la neutralidad. Esta puede deberse a temor al compromiso de tomar posición, a inseguridad en relación con la respuesta (y se escoge la neutral para equivocarse menos) o a que, en realidad, hay una posición neutral. El investigador quizá quiera saber si en algún tema la opinión predominante es la neutral.

Si el número de opciones es par, eso permitirá identificar si hay una tendencia hacia uno de los extremos, ya que se obliga a los encuestados, de cierta manera, a tomar una posición diferente de la neutral.

EJEMPLO

¿La investigación requiere creatividad?

Totalmente de acuerdo	De acuerdo	Neutro o sin opinión	En desacuerdo	Totalmente en desacuerdo

Para obtener una mayor diversidad, en algunas ocasiones se ponen siete espacios o más de respuesta, de acuerdo con la pregunta. Se recomienda utilizar escalas de cinco puntos para escalas unipolares (por ejemplo, escalas que oscilan desde "extremadamente enfermo" a "para nada enfermo"), y de siete puntos, para escalas bipolares (por ejemplo, de "extremadamente enfermo" a "extremadamente sano").

Escala diferencial semántica (o escala de Osgood)

Su fin es obtener respuestas a una serie de dimensiones acerca de un objeto o una categoría determinados. El supuesto fundamental de este tipo de escalas es que el resultado de una evaluación o un juicio puede concebirse como el lugar que el objeto evaluado ocupa en un continuo definido por dos términos (***adjetivos***) **bipolares.**

EJEMPLO

Su depresión es…

Dolorosa	No dolorosa
Vergonzosa	No vergonzosa
Grave	Leve
Incapacitante	No incapacitante

Además de las respuestas escalares referidas, hay otro tipo de métodos, como los comparativos: el método de Thurstone, la técnica de comparación pareada y el método de Gutman, así como métodos econométricos (la apuesta estándar y la técnica del intercambio por tiempo[11-14], todos los cuales, por salirse de la temática del presente capítulo, no serán revisados.

CONCLUSIÓN

En este capítulo hemos revisado los pasos fundamentales para la elaboración de cuestionarios. El primer paso es formular y refinar cuidadosamente las hipótesis que se están evaluando, decidir la información que quiere recolectarse y revisar las referencias científicas que lo sustenten. Después hay que elegir el tipo de cuestionario y el tipo de pregunta por cada variable, y cómo va a codificarse cada respuesta. Después hay que diseñar el formato, preparar manuales de instrumentos y hacer un piloto para revisar el cuestionario.

Es fundamental que las preguntas, ya sean abiertas o cerradas, estén bien planteadas, y que se ponga un énfasis especial a su redacción y su organización. Siguiendo estos pasos, es posible obtener información sobre actitudes, creencias, comportamientos y percepciones, que pueden ser de gran utilidad en estudios epidemiológicos.

Referencias

1. Sánchez R, Gómez C. Conceptos básicos sobre validación de escalas. Rev. Col. Psiquiatría, 1998;27:121-30.
2. Knapp RG, Clinton M. Clinical epidemiology and biostatistics. Baltimore: Williams and Wilkins; 1992.
3. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: The essentials. 3rd ed. Baltimore: Williams and Wilkins; 1996.
4. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38:27-36.
5. Phillips MR, Yangping Z. Methodological framework for the evaluation and revision of translated instruments. En: Clinical applied health social science. Australia: The University of Newcastle; 1995.
6. Streiner DL. Scale construction checklist. Mc Master University Document. Unpublished notes.

7. DeVellis RH. Scale development. Theory and applications. Newbury Park: Sage Publications; 1991.
8. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. New York: Oxford University Press; 1995.
9. Streiner DL. A checklist for evaluating the usefulness of rating scales. Can J Psychiatry Rev Can Psychiatr. 1993;38:140-8.
10. Spielberger CD, Sharma S. Cross-cultural measurement of anxiety. En: Phillips MR, Yangping Z. Methodological framework for evaluation and revision of translated instruments. Integrating Health Social Science in the Design & Assessment of Clinical Research Protocols. Australia: University of Newcastle; 1995.
11. Phillip MR. Evaluation and revision of translated western scales. En: Wang XD. Rating Scales for mental health. Beijing: Chinese J Mental Health. 1993; 343-51. (Unpublished translation).
12. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. Nueva York: John Wiley & Sons; 1981.
13. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40:171-8.
14. Guyatt GH, Berman LB, Townsend M, et al. Should study subjects see their previous responses? J Chronic Dis. 1985;38:1003-7.

CAPÍTULO

29

Plan de análisis

María Nelcy Rodríguez Malagón
Juan Gabriel Ruiz Peláez

INTRODUCCIÓN

El plan de análisis es un componente indispensable en la elaboración de cualquier proyecto de investigación. Los análisis deben planearse cuidadosamente de antemano, deben ser la consecuencia lógica de la formulación de la pregunta de investigación y deben articularse con las consideraciones acerca de población y muestra, con la enumeración y la definición de las variables de interés y su medición, y, especialmente, con la estrategia de investigación escogida. El plan de análisis es una carta de navegación racional que prevé y anticipa la estrategia de recolección y tratamiento de la información y plantea las descripciones, las evaluaciones, los contrastes y los ajustes que se llevarán a cabo para contestar adecuadamente las preguntas de investigación.

Hacer una investigación sin definir previamente un plan de análisis explícito es totalmente inapropiado, por dos razones:

- Sin un plan de análisis, la recolección de datos es heterogénea (por no decir caótica), no va dirigida a un fin específico y es muy difícil asegurar su calidad y la correcta administración y protección de los datos obtenidos.
- "Jugar" con los datos recolectados para ver cuáles asociaciones se pueden establecer puede generar asociaciones reales, pero también, otras espurias; estas últimas, sin racionalidad ni plausibilidad. Tales ejercicios de talante inductivo, en vez de producir conocimiento científico, producen ruido y confusión.

Formular y evaluar hipótesis ***post hoc*** en un mismo ejercicio de investigación es riesgoso. Las asociaciones aparentes que surgen después de examinar los datos recolectados son útiles como fuente de hipótesis o de conjeturas, pero jamás como fuente primaria de evidencia de "hechos" científicos. Es decir, las conjeturas generadas ***post hoc*****, solo si son plausibles y** relevantes, deben ser evaluadas en ejercicios de investigación subsecuentes, en los cuales la metodología, la recolección de datos y el análisis están guiados por esas nuevas hipótesis.

En síntesis, la práctica de "improvisar" los análisis después de recogidos los datos es una invitación a forzar las conclusiones y a imponer interpretaciones sesgadas de los datos, lo cual se constituye en investigación de pobre calidad y es cuestionable, incluso, éticamente.

Ahora bien, es apropiado reconocer que un plan de análisis es eso mismo: un plan. Es necesario (de hecho, un buen plan lo contempla) ajustar las estrategias y las tácticas empleadas a

la realidad de la conducción del proyecto de investigación, pero sin perder la dirección general de los procesos, planteados desde la formulación del protocolo de investigación.

Un buen plan de análisis contempla aspectos a lo largo de todas las fases del proyecto, y no es, como podría juzgarse equivocadamente, una etapa aislada, al final de una investigación y a cargo de alguien que no ha participado en la planeación ni el desarrollo del proyecto. La idea (bastante difundida entre investigadores sin una sólida formación metodológica y bioestadística) de que una vez "realizado" el estudio y recogidos los datos se puede vincular a alguien con experiencia en metodología y bioestadística para que "haga las pruebas y calcule los valores de *p*" es totalmente inapropiada. Esta especie de subcontratación del análisis es también una forma de improvisación del análisis, y, por tanto, una mala práctica científica y ética.

Muy al contrario, la planeación y la conducción del análisis de una investigación son actividades integradoras que implican una interacción entre todos los aspectos metodológicos y procedimentales del estudio. Las estrategias de análisis, las pruebas y los contrastes escogidos deben responder, entre otros aspectos, a los siguientes componentes de la investigación:

- Tipo de pregunta planteada.
- Objetivos propuestos.
- Estrategia de investigación o diseño escogido.
- Definición de la población.
- Definición de la muestra, incluido su tamaño.
- Técnicas de selección de la muestra que se hayan definido.
- Definición y clasificación de las variables o los eventos de interés que serán medidos.
- Métodos de medición empleados para cuantificar o clasificar las variables de interés y su capacidad de discriminación (sensibilidad y especificidad; confiabilidad).
- Número de grupos de comparación, si es el caso.
- Número de observaciones que se tomarán por sujeto.
- Proceso de recolección de datos; confiabilidad y precisión de la recolección; procesamiento y registro de la información (respuestas de los participantes, diarios, etc., *vs*. mediciones bioquímicas, evaluación de imágenes diagnósticas, etc.).
- Selección de las técnicas estadísticas apropiadas para el procesamiento, la preparación y las transformaciones de los datos.
- Supuestos necesarios y establecidos para decidir sobre el análisis estadístico de la información recolectada.

El plan de análisis debe responder al objetivo principal del estudio, que es, en último término, buscar respuesta a las preguntas de investigación.

El desarrollo apropiado del análisis contempla el desarrollo de pasos (a veces, en etapas sucesivas, y algunas, en paralelo) hasta tener las respuestas, incluyendo aspectos numéricos y cuantitativos, a las preguntas de investigación **(tabla 29.1)**.

Identificar estos pasos facilita el proceso de planeación temprana y ordenada que especifique, de forma precisa y detallada, cómo se conducirá el análisis de los datos. Dicha planeación está condicionada, entre otras cosas, por el diseño de los instrumentos o los formatos de recolección de datos, con el fin de garantizar que la información que se quiere obtener se consigne de forma completa y con los grados de validez y de precisión necesarios. Una vez recolectada la información, debe procesarse, verificar que es correcta y completa, modelarla para que sea útil, y, algunas veces, transformarla para que se facilite el análisis.

Aun cuando el proceso de aseguramiento de la calidad de los datos se hace de forma continua durante el seguimiento de los sujetos y la captura de la información, el momento en el que se completan la recolección y el registro de la información es un hito especial en el curso de la investigación. En ese momento, es apropiado hacer una primera revisión general de la base de datos ya completa, que permita hacer una última edición y la verificación sobre su calidad, que facilite la toma de decisiones finales sobre las pruebas estadísticas que deben emplearse según los objetivos y las características de los datos[1], y, en general, que dé una idea de lo que se ha en-

[1] Vale la pena recordar que las estrategias planeadas de antemano definieron ***a priori*** los tipos de pruebas y contrastes para realizar, pero que los ajustes finales (sobre todo, la verificación empírica de que se cumplan los supuestos necesarios para la utilización de determinadas pruebas estadísticas) solo se pueden hacer con los datos ya recolectados.

Tabla 29.1 Pasos en el análisis de datos en un estudio

1	Plan de análisis explícito, *a priori*.
2	Recolección estandarizada de información (incluye control de calidad en todas sus etapas).
3	Limpieza y procesamiento de la información.
4	Descripción y caracterización de la población o de la muestra estudiada, incluyendo la distribución de las variables medidas y la evaluación de supuestos necesarios para escoger indicadores y pruebas y análisis exploratorio.
5	Análisis inicial, o "simple" (incluye estimación de frecuencias, o de efectos "crudos").
6	• Análisis confirmatorio. • Evaluación de la necesidad de ajuste. • Estimaciones ajustadas de frecuencias o de efectos.

contrado y de cómo debe analizarse. En últimas, el análisis propiamente dicho confirma o refuta los datos preliminares, solidifica las conclusiones y sienta las bases para la generalización científica o estadística de los resultados a la población de interés, bien sea esta una ***población*** *de referencia* (abstracta) o una ***población blanco*** *específica* (concreta).

PLAN DE ANÁLISIS *A PRIORI*

El planteamiento de una pregunta de investigación clara y explícita permite escoger y definir las estrategias más apropiadas para conducir el análisis final de la información. Las estrategias cambian según el tipo de objetivo general del estudio: hacer estimaciones (***descripciones***) o hacer evaluación de hipótesis (***analíticas***). También influyen otros aspectos metodológicos, tales como la definición y la medición de las variables de interés, la determinación de una población de estudio y el proceso de recolección de información de acuerdo con el tipo de diseño propuesto. A continuación se revisan algunos de estos aspectos.

Objetivos del estudio

Un correcto planteamiento de los objetivos tanto primarios como secundarios (si los hubiera) de un estudio constituye el primer paso para establecer ***a priori*** la estrategia apropiada de análisis.

Si el objetivo del estudio es descriptivo (por ejemplo, determinar la frecuencia de complicaciones pulmonares en trabajadores expuestos a partículas de asbesto), a partir de dicho objetivo se define cuál es el evento o la variable de interés en el estudio (incidencia o prevalencia de la enfermedad pulmonar), cómo se va a medir (pruebas de función pulmonar, reporte de los sujetos, informes de autopsias, etc.); también se define el tipo de sujetos a quienes se debe incluir para que se cumpla el objetivo (trabajadores expuestos al asbesto), otros criterios de inclusión y de exclusión (por ejemplo, la ausencia de comorbilidad pulmonar) y el intervalo a lo largo del cual se deben observar, y se podrán escoger de antemano los estadígrafos y las técnicas estadísticas apropiadas.

Si, en cambio, el objetivo primario del estudio es analítico (como, por ejemplo, establecer si hay alguna diferencia en la frecuencia de fibrosis pulmonar entre trabajadores expuestos y no expuestos a partículas de asbesto), el proceso de análisis varía, debido al cambio en los objetivos y el consecuente ajuste en el tipo de estrategia de investigación. En este caso, el análisis compara, de forma tanto ***cruda*** (primaria) como ***ajustada*** (si es necesario el ajuste), la incidencia de la fibrosis pulmonar, y estima la presencia, la dirección y la magnitud de la asociación entre las exposiciones y los desenlaces.

Hipótesis conceptual e hipótesis estadística

Como se discutió en el **capítulo 2**, "La pregunta de investigación", en investigación cuantitativa hay dos tipos básicos de preguntas: las llamadas ***descriptivas***, cuya respuesta es una descripción o un mapa de la realidad, y las ***analíticas***, que indagan sobre la asociación entre factores y desenlaces, y cuyas respuestas son una predicción de la ocurrencia de eventos o, mejor aún, una explicación a por qué y cómo ocurren los eventos.

Preguntas descriptivas

En principio, los diseños de investigación descriptiva no evalúan hipótesis conceptuales de asociación, sino que buscan medir o estimar características en conjuntos de sujetos definidos por alguna característica (es decir, en poblaciones definidas). Tal como se explicó en el **capítulo 5**,

"Población y muestra", se estudian muestras representativas de poblaciones de interés (*poblaciones blanco*), para evaluar, mediante el estudio de muestras, los parámetros de interés en dichas poblaciones. Una forma de enfocar el análisis de tales estudios descriptivos consiste en plantear hipótesis estadísticas ***descriptivas***, en las cuales la hipótesis nula (que se quiere intentar refutar) es que el valor del estimador no es diferente del valor de un supuesto parámetro poblacional esperado, e, incluso, que el estimador no es diferente de 0.

La prueba de hipótesis se hace estimando la probabilidad (el valor de "p") de que las diferencias entre el estimador y el hipotético parámetro poblacional sean reales, o, por el contrario, producto de la variación de muestreo. Si la probabilidad estimada "p" de que las diferencias observadas en la muestra no sean reales y obedezcan a variación de muestreo queda por debajo de un valor crítico (un umbral de significación; por ejemplo, 0,05), se dice que la diferencia es ***estadísticamente significativa***, y se dice que con los datos provenientes de la muestra hay evidencia para rechazar la hipótesis nula de no diferencia.

Este test de hipótesis en el escenario descrito es un tanto forzado y artificial, y su interpretación no es evidente. Puede resultar más claro y eficiente prescindir por completo del planteamiento de hipótesis estadísticas descriptivas, y enfocar, más bien, el análisis en la precisión de las estimaciones, medidas en la forma de intervalos de confianza (IC) en torno al estimador puntual.

Preguntas analíticas

El planteamiento de hipótesis tanto conceptuales como estadísticas es de verdad relevante en la investigación analítica (predecir o explicar, evaluar asociación entre exposiciones y desenlaces).

En el caso de las preguntas analíticas, el objetivo principal del estudio es contrastar valores de variables de desenlace entre grupos definidos por su nivel de exposición a una o más variables independientes de interés. La pregunta de investigación analítica, según lo explicado en el **capítulo 2 "La pregunta de investigación"**, puede ser manejada de forma más eficiente si se plantea como un enunciado o una conjetura científica (hipótesis) que hace predicciones empíricas evaluables, y que es susceptible de ser refutada con la información proveniente de los datos recolectados.

La especificación de las predicciones hechas por la hipótesis conceptual (los hallazgos esperados si, y solo si, la hipótesis fuera correcta) en términos de ***magnitudes*** (valores numéricos) de los resultados esperados, permite plantear hipótesis estadísticas que pueden ser examinadas mediante pruebas estadísticas específicas.

Debido a la falacia de la inducción ("no se puede hacer inferencia lógica sobre principios generales, a partir de observaciones particulares"), los datos empíricos no pueden confirmar una hipótesis. Pero los datos empíricos sí pueden rechazar una hipótesis, cuando se establece que la información empírica contradice las predicciones hechas por la hipótesis. La forma más eficiente de hacer tal tipo de pruebas estadísticas consiste en plantear como ***hipótesis nula*** (que se quiere falsar) la afirmación de que los estimadores obtenidos en los grupos expuestos a diferentes niveles de la variable independiente de interés no son "estadísticamente diferentes" entre sí (con un nivel dado de precisión de estimación, o de ***confianza*** en la estimación), o, mejor aún, no son diferentes de un valor esperado en una población de la cual provienen los sujetos (un ***parámetro poblacional común***).

Esta hipótesis compite con la explicación alternativa (***hipótesis alternativa***) de que las diferencias halladas en los valores estimados de la variable de desenlace en cada grupo no se deban a variación de muestreo, y que, en realidad, los sujetos de los diferentes grupos no provengan de una misma población con un determinado parámetro poblacional, sino que en razón de la exposición diferencial ahora provienen de poblaciones distintas, con valores paramétricos de la variable de desenlace que son diferentes entre sí. El rechazo de la hipótesis nula no confirma la hipótesis alternativa: solo la deja como la explicación más verosímil para los hallazgos.

Tipo de estudio

La estrategia de investigación seleccionada para responder a la pregunta de investigación determinará, en gran medida y desde la fase de diseño, cuál será el plan de análisis. Si el estudio es eminentemente descriptivo (como, por ejemplo, una serie de casos) o un estudio de prevalencia o de tipo transversal (donde solo se busca precisar con qué frecuencia ocurre un

evento), la metodología de análisis será diferente de la seleccionada cuando el estudio sea de tipo analítico, donde el objetivo es siempre la evaluación de hipótesis. El primer caso supondrá utilizar solo técnicas estadísticas descriptivas, mientras que en el segundo, además de dichas técnicas, se debe considerar el uso de técnicas de inferencia estadística para evaluar hipótesis en la comparación de grupos.

Por ejemplo, si el estudio pretende establecer si hay alguna asociación entre el grado de estrés y los intentos de suicidio entre los jóvenes de cierta comunidad, durante la fase de análisis de datos se debe hacer no solo un análisis descriptivo inicial de la población estudiada y de cada una de las características medidas, sino también el cálculo y la interpretación de índices especiales (llamados *medidas de asociación*; véase el capítulo 9 "Medidas de frecuencia de asociación y de impacto") que permitan evaluar si realmente hay evidencia de asociación y si dicha asociación es afectada por otros factores que pueden introducir modificación de efecto (***interacción***) o confusión de la supuesta asociación (***factores de confusión***).

Determinar si la supuesta asociación encontrada está confundida o modificada por factores extraños, diferentes de la exposición de interés, es un aspecto crucial del análisis de un estudio que busca evaluar asociación. En caso de hallar evidencia de confusión o modificación de efecto, es necesario hacer ajustes para corregir la estimación afectada. Se emplean diferentes estrategias, tan sencillas como la estratificación del análisis según los niveles del potencial factor de confusión, hasta la generación de modelos matemáticos complejos (***análisis multivariados***).

EJEMPLO

Si se evalúa la frecuencia de infarto del miocardio en una población expuesta a distintos grados de tabaquismo, la edad es también un factor que influye en la aparición del infarto. Si, adicionalmente, la intensidad del tabaquismo cambia con la edad, dicha variable (edad), al estar asociada a la exposición (al tabaco; a mayor edad, mayor exposición) y al desenlace (a mayor edad, más posibilidad de infarto) podría confundir cualquier asociación aparente entre consumo de cigarrillo e infarto del miocardio. En el análisis se debe separar (*estratificar*) a los pacientes por grupos etarios, para controlar la posible modificación de efecto introducida por esta variable y dejar así al descubierto el verdadero impacto del cigarrillo.

Cuando se evalúa la modificación de efecto introducida por la edad en la asociación entre tabaco e infarto, puede descubrirse que afecta de manera uniforme a todos los estratos (***confusión***) o que afecta diferencialmente a algunos de los estratos (***interacción***). Los conceptos de confusión e interacción serán abordados de forma más extensa en este mismo capítulo.

Población de estudio y diseño de la muestra

Un plan de análisis establecido ***a priori*** debe tener en cuenta si los sujetos de estudio son todos los posibles (en este caso, hablamos de un estudio por censo) o si, por el contrario, el grupo investigado es solo una parte o una fracción de toda la población.

El hecho de adelantar un estudio basándose en una muestra supone considerar algunos aspectos:

- Los objetivos mismos del estudio.
- La variabilidad de los sujetos respecto a la variable de interés.
- El cálculo del número de sujetos que debe ser estudiado; a su vez, dicho cálculo está determinado por varios factores:
  - *Error de tipo I* (error α), o aceptación errónea de una diferencia que no existe.
  - *Error de tipo II* (error β), o negación errónea de una diferencia que no se pudo detectar, pero realmente existe.
  - Poder del estudio que se relaciona con el error β.

La diferencia mínima que se quiere determinar; especialmente, cuando se incluye una comparación entre grupos.

Los mencionados factores son relevantes para estudios que evalúan hipótesis; es decir, comparaciones entre dos o más grupos. Cuando se busca estimar un parámetro poblacional, las consideraciones de tamaño de muestra son diferentes: la variabilidad de la característica en la población blanco y la precisión deseada en la estimación del parámetro (anchura del IC en torno al cálculo del estimativo puntual).

Si el estudio ha empleado una muestra de sujetos, entonces el análisis ***a priori*** debe establecer técnicas de inferencia estadística cuando se analicen los resultados obtenidos, aunque se trate de un solo grupo de estudio o de la comparación de dos o más grupos.

Definición y enumeración de variables

Las *variables* son las características que varían de un individuo a otro, y cuya medición da como resultado un tipo de dato específico. La definición de variables o de eventos de interés en un estudio es otro de los factores que determinan el tipo de análisis que será propuesto; según el tipo y el número de variables que se definan, se podrá determinar si el análisis del estudio propuesto es de tipo *cualitativo*, de tipo *cuantitativo* o una combinación de estos, al igual que si el análisis será univariado, bivariado o multivariado[1-3].

Un aspecto muy importante a la hora de establecer un análisis *a priori* sobre la información que será recolectada en un estudio es el tipo de escala de medición que se haya empleado en cada una de las variables. Si estas son eminentemente de tipo cualitativo o medidas en una *escala de intervalo*, la técnica de análisis estadístico será diferente de la seleccionada si las variables medidas son eminentemente cuantitativas y medidas en una *escala de razón*.

Si el estudio involucra medir un conjunto de variables cualitativas y cuantitativas, y, además, se busca establecer posibles asociaciones entre varias de ellas, el análisis de datos deberá considerar este hecho[4]. Por último, si la investigación también pretende establecer la presencia de una o más variables de confusión o de interacción (*modificadores de efecto*), igualmente, el análisis de información que se determine *a priori* debe tener en cuenta dicha posibilidad.

Por otro lado, algunos estudios consideran solo la descripción de un evento, sin tener en cuenta los otros. En ese caso, el análisis usará solo técnicas estadísticas de tipo descriptivo para una sola variable (distribuciones de frecuencia simple o medidas de tendencia central y de variabilidad). En otros casos, se debe evaluar no solo la variable principal (o *dependiente*), sino también, todos los otros factores asociados, o *variables independientes*, pues en tal circunstancia el plan de análisis que se establezca a priori tendrá en cuenta que debe hacerse un análisis multivariado y preverse el uso de las técnicas de análisis apropiadas[5].

Una vez decidido cuál será el tipo de análisis, conviene diseñar el esquema completo de métodos para presentación de datos, mediante las denominadas *tablas vacías* (*dummy tables*) —esquema de lo que, finalmente, se quiere obtener al terminar el proceso de recolección y análisis de datos[6]—.

RECOLECCIÓN DE INFORMACIÓN

El proceso que se utilizará para la recolectar información es otro de los aspectos que determinan el tipo de análisis que debe hacerse. Por ejemplo, si el estudio es una encuesta, la información debe recolectarse mediante formatos o cuestionarios previamente estandarizados; o sea, verificados por medio de una prueba piloto o de algún otro sistema de comprobación sobre la claridad en el planteamiento de las preguntas involucradas, el orden de los temas que se abordarán, el lenguaje utilizado, el flujo y la longitud de las preguntas y la codificación de las alternativas de respuesta para las preguntas de tipo cerrado[7].

En este caso, el plan de análisis que se fije *a priori* debe considerar elementos como el número de encuestas, el porcentaje de estas que sean respondidas por completo, el número que se puede aceptar que se pierda (por falta de participación en el estudio, por información incompleta, etc.). Si el estudio es un experimento clínico, recolectar información supone el uso de formatos que serán llenados tanto por encuesta como por observación directa, algunas veces. Ante tal circunstancia, debe planearse con antelación si se harán o no comparaciones múltiples, cuántas de ellas, cuándo (*análisis interino*) y cómo se hará, finalmente, el análisis (evaluaciones de subgrupos y utilizando o no la intención de tratar).

Procesamiento y limpieza de la información

El procesamiento y la limpieza de la información implican un proceso continuo que puede afectar de forma positiva un plan de análisis de datos, pero también puede hacerlo de forma muy negativa. Un plan de análisis definido preliminarmente se elabora con los supuestos de una buena calidad de los datos, calidad que en ocasiones se ve afectada por alteraciones en el plan de recolección de datos o de su procesamiento. Antes de su análisis, los datos se deben recolectar mediante un proceso de excelente calidad; deben ser luego revisados, codificados quizá —si dicho aspecto no se consideró durante

la fase de diseño—, procesados y convertidos de una forma tal que puedan ser finalmente analizados[8].

Recolección de datos

Decimos que recolectar datos es un proceso continuo porque durante el transcurso de la recolección de datos se deben vigilar ciertos aspectos[3,9,10]:

Control de calidad

Se refiere a la verificación de los datos consignados, la corrección de los que se consideren inconsistentes, la repetición de un porcentaje de las entrevistas (o de algunas en particular) para verificar la confiabilidad de la información, la doble revisión sobre la información faltante, etc. En el caso de los estudios experimentales, se revisa, además, la elegibilidad de los pacientes para el estudio, se verifica su asignación aleatorizada, se asegura la consignación adecuada de la información en los formatos, se garantiza el seguimiento adecuado de los pacientes y se revisa que los formatos de recolección hayan sido llenados adecuadamente. Cuando se trate de un estudio multicéntrico, el control de calidad de cada aspecto debe vigilarse cuidadosamente en todos los centros.

Consistencia y validez de las mediciones

Aunque es un aspecto incluido en el control de calidad, vale la pena mencionarlo como un numeral aparte, pues cobra mayor importancia, especialmente, en los estudios experimentales. Los instrumentos de medición deben estar debidamente estandarizados y calibrados para garantizar tanto la confiabilidad como la validez de los datos obtenidos. En un estudio de tipo transversal donde se emplee una encuesta, los encuestadores deben estar debidamente estandarizados en todo el proceso de aplicación de formatos de recolección, y en los demás procesos para recolectar información en campo que sean pertinentes al estudio.

Codificación y poscodificación

Si bien algunos de los datos tienen formatos precisos y concretos (fecha de nacimiento, peso, talla), algunos de estos deben clasificarse (estado de la hipertensión arterial, grado de sobrepeso) o deben ser codificados para facilitar la manipulación o el análisis (sexo masculino = 1 y femenino = 2; edad < 15 años = 1, entre 15 y 45 años = 2, y mayor de 45 = 3).

Habitualmente, al preparar los formatos de recolección se hace la clasificación (***precodificación***), y el encuestador o el monitor la usan durante el registro de los datos. En algunos casos, debe clasificarse después de recolectarse la información (***poscodificación***).

Las codificaciones pueden influir en el análisis, bien sea facilitándolo (análisis según peso normal = 1, sobrepeso = 2 y obesidad = 3), bien sea mejorando la posibilidad de manipular la información (análisis por sexo = 1 o 2; control por tabaquismo = 0 o 1) o bien sea limitando la información global para facilitar un análisis (reducción de la edad, variable continua, a 3 categorías (1 = < 15 años; 2 = 15-45 años; 3 = > 45 años).

Procesamiento de la información

Cualquier plan de análisis definido ***a priori*** puede, igualmente, verse alterado por un procesamiento de datos no conducido según lo planeado. Esto puede ocurrir en alguno de los pasos que componen el proceso: en la falta de un manual de códigos o en el de entrada y verificación de datos[3,10].

Manual de códigos

El procesamiento de datos considera, como primera medida, contar con un diccionario de variables o un manual de códigos donde se incluyan no solo los nombres de las variables con sus respectivos códigos, sino las que son definidas o calculadas a partir de las inicialmente medidas, las que son recodificadas para el análisis final y las que han sido transformadas de su escala de medición original (por ejemplo, una variable que es cambiada a una escala logarítmica), para efectos de análisis.

Proceso de entrada y verificación de datos

En algunos estudios se puede considerar la aplicación de un plan de análisis cuyos datos han sido procesados manualmente, pero eso significa que no se ha incluido a muchos sujetos en el estudio ni se han hecho muchas mediciones sobre ellos. Sin embargo, en la mayoría de los casos, es necesario recurrir al uso de bases de datos y de programas que permitan llevar a cabo tanto el proceso de entrada de datos como la verificación de sus inconsistencias.

Antes de introducir la información en una base de datos, se supone que se la ha verificado; a pesar del paso por este filtro, no es raro que vuelvan a aparecer inconsistencias. En tales circunstancias, se puede verificar de nuevo, mediante el ***sistema de doble digitación.***

ANÁLISIS EXPLORATORIO INICIAL

Su objetivo es ser un paso más en el proceso de edición de datos; o sea, se pueden hacer distribuciones de frecuencias simples de todas las variables incluidas en el estudio, en las cuales es fácil detectar errores que se pasaron por alto durante alguna de las fases de recolección o de procesamiento de la información. Por ejemplo, en una distribución simple de frecuencias puede detectarse que hay valores extremos anormales (talla = 2,35 m; peso = 1135 kg), valores ilógicos (sexo = 3) o fechas erradas (nacimiento 14.14.1987).

Detectar valores extremos o errores en la codificación de alguna de las variables debe buscar la corrección no solo del error, sino de la situación que lo generó, lo cual significa, ocasionalmente, revisar la fuente primaria, repetir una medición, verificar los criterios de inclusión, o hasta eliminar algunas observaciones.

Adicionalmente a la utilidad como paso final en el proceso de edición de los datos, el análisis exploratorio sirve a la hora de determinar las técnicas estadísticas finales que deban emplearse, una vez conocidas las características de los datos de los cuales se dispone. Por ejemplo, en la exploración inicial se encuentra que una variable no tiene una ***distribución normal***, lo cual exige un análisis especial.

En esta fase del estudio se considera el uso de técnicas estadísticas que, básicamente, se aplican para describir datos[2,3]. Vale la pena aclarar que los factores considerados determinantes en la selección de la técnica estadística que será aplicada en un estudio son[2]:

- Los objetivos y el propósito del estudio.
- Las características matemáticas de las variables de interés.
- Los supuestos estadísticos sobre las variables que se han medido.
- El proceso de recolección de los datos.

Estos cuatro factores determinan si el análisis exploratorio debe considerar solo la obtención de distribuciones de frecuencias simples para cada una de las variables de interés que fueron medidas, o si, adicionalmente, será necesario buscar medidas de tendencia central y de variabilidad, o de dispersión u otras medidas (***medidas*** *de precisión*).

Según los resultados, también se puede determinar si se tiene un tamaño de muestra suficiente para cumplir con el plan de análisis propuesto ***a priori***. La selección de las técnicas de análisis exploratorio de datos está, igualmente, determinada por[11]:

- El tipo de datos recolectado: cualitativos o cuantitativos.
- La estrategia de diseño utilizada.
- El número de grupos estudiados.
- El tamaño final de la muestra.

Si se trata de dos o más grupos, su grado de relación o independencia (muestras independientes o apareadas).

ANÁLISIS FINAL

En esta fase se considera la aplicación de técnicas estadísticas que ya han sido determinadas por todas y cada una de las fases anteriores del proceso. El plan de análisis se aplicará total o parcialmente según los objetivos, el tipo de diseño del estudio, el número de variables y el cumplimiento estadístico de los supuestos matemáticos sobre estas, el tipo de población de estudio, la muestra seleccionada, el número de grupos estudiados, etc.

Según las características de las variables de interés y de los supuestos matemáticos que se adopten, se decidirá si se hace un análisis univariado o multivariado, si se utilizan técnicas paramétricas por estar en presencia de variables con distribución normal o si se hace necesario el uso de pruebas no paramétricas[10,12,13].

Mediciones crudas (análisis "simple")

Un primer paso en el análisis final es calcular las medidas que pueden describir la ocurrencia de un evento, y que son las que solo presentan la frecuencia tal ocurrencia. Un ejemplo de dichas medidas son las ***tasas de prevalencia simple***, en el caso de los estudios de corte transversal; o las

tasas de incidencia acumulada, en el caso de estudios de cohorte[14]. La fase del análisis dentro del cual se calculan esas medidas se conoce también como ***análisis simple***[15].

Exploración de confusión e interacción, y mediciones ajustadas

La ocurrencia de un evento de interés puede verse afectada por la presencia de uno o de varios factores. Ello hace necesario que la descripción sobre tal evento deba ser "ajustada" de alguna forma por los efectos de dichos factores. Por ejemplo, al comparar dos poblaciones respecto a una tasa de mortalidad global, puede ser necesario el ajuste de dicha medida por la distribución diferencial de la edad entre esas poblaciones. Ello significa que la tasa de mortalidad cruda puede ser muy diferente para cada población, respecto a la obtenida cuando se hace el ajuste por la distribución de la edad en cada una las dos poblaciones.

Pasos

Estimación cruda de efectos

Como ya se mencionó, cuando se hace el plan de análisis para un estudio, tanto los objetivos, el diseño o la estrategia de estudio como la medición de las variables de interés determinan el tipo de medidas de tipo estadístico que se obtendrán. Si el estudio, además, no solo considera la descripción de un evento, sino la posible asociación entre la ocurrencia de dicho evento y uno o más factores independientes, se debe, entonces, identificar el tipo de medida o de medidas estadísticas que permitirán evaluar el efecto de esos factores sobre el evento de interés. Todo ello implica conocer la magnitud de la asociación cruda. La forma de cálculo de dicha asociación varía según la estrategia de diseño del estudio. Así, por ejemplo, en un diseño de cohortes, el riesgo relativo (RR) es la medida de elección, mientras que en un diseño de casos y controles (CC), el riesgo relativo indirecto (*odds-ratio*) es la medida que refleja la magnitud de la asociación en estudio.

Evaluación de modificación de efecto

La evaluación de la asociación entre dos o más variables puede estar influida por otros factores. Si ese es el caso, la estimación cruda del efecto no reflejaría efectivamente tal asociación. Dos aspectos importantes se deben evaluar en tales circunstancias son: 1) el efecto de la confusión y 2) el efecto de la interacción.

Confusión

El efecto de confusión es evidente cuando se obtiene un resultado diferente de la asociación entre dos variables, por la presencia o no de una tercera variable[15]. Tal diferencia se percibe al calcular las medidas de efecto cruda y ajustada. Por ejemplo, en el caso de un estudio de cohorte para evaluar la asociación entre un nuevo tratamiento y la reducción en los niveles de tensión arterial, la magnitud del RR puede ser diferente cuando se tiene en cuenta (o se ajusta) por la presencia de un tercer factor, como el consumo de alcohol.

Un plan de análisis debe considerar, entonces, la posible presencia de variables de confusión al evaluar la asociación o la relación de interés. Quiere esto decir que dicha variable estaría modificando la asociación que se busca evaluar.

Interacción

Al igual que en el caso de la confusión, la relación entre dos o más variables puede verse modificada a diferentes niveles de una tercera variable. Ello se da cuando se evalúa la relación de interés y la medida de asociación seleccionada varía tanto en magnitud como en cuanto a la interpretación clínica, a diferentes niveles (o *subgrupos*) de la variable edad. Si se presenta dicha variación, entonces será necesario que el análisis final de estudio considere la evaluación sobre la presencia potencial de una interacción. De nuevo, una medida de la asociación cruda no reflejaría la asociación que un estudio pretende evaluar. En el ejemplo sobre la asociación entre el nuevo tratamiento y la reducción de los niveles de tensión arterial, si dicha reducción es diferente cuando se tienen en cuenta distintos grupos de edad, diremos que hay una interacción entre el tratamiento y la edad.

Estimaciones ajustadas de efecto

Ante la posible presencia de interacción, o de ***efecto de confusión***, un plan de análisis debe considerar cuál será la estrategia para evaluar la asociación en tales circunstancias.

Ello implica la posibilidad de usar herramientas tales como el análisis estratificado o

el uso de modelos matemáticos. La estadística ofrece técnicas cada vez más sofisticadas en tal tipo de herramientas. Los modelos matemáticos, como los de tipo loglineal, de regresión logística y de supervivencia, han tenido un desarrollo muy amplio a lo largo de los últimos años, y le dan al investigador la posibilidad de ajustar la relación entre el evento de salud que le interese por más de un factor independiente de forma simultánea.

CONCLUSIÓN

El plan de análisis de un estudio debe ser planteado a la luz de aspectos como la pregunta de investigación, los objetivos del estudio, la medición de las variables y el tipo de datos obtenidos de su observación, provenientes, a su vez, del estudio de toda la población (***censo***) o de una parte de ella (***muestra probabilística*** o *muestra de conveniencia*), la estrategia de diseño empleada y la forma de recolección de los datos.

El planteamiento claro de un primer análisis, o ***análisis exploratorio***, le permitirá al investigador tener guías o pautas sobre la obtención de los primeros resultados de su estudio y sobre las técnicas estadísticas que pueden ser aplicadas en el análisis final de la información. Dicho análisis final debe considerar no solo la estimación y el cálculo de medidas crudas o simples, sino el posible ajuste por variables que pueden modificar no solo la descripción sobre la ocurrencia de un evento, sino el estudio de su posible asociación a uno o más factores de exposición. La estadística le provee al investigador una variedad de técnicas que le permiten todos esos tipos de análisis.

Referencias

1. Norman GR, Streiner DL. Biostatistics. The Bare Essentials. Philadelphia: Mosby; 1994.
2. Taylor C, Gibbon F, Morris L. How to analyze data. California: SAGE Publications; 1987.
3. Hulley SB, Cummings SR. Designing clinical research. An epidemiologic approach. Baltimore: Williams & Wilkins; 1988.
4. Roy DJ, McL P, Black B, et al. Ethical principles in research. En: Troidl H, Spitzer WO, McPeek B, H, et al. Principles and practice of research. Nueva York: Sringer-Verlag; 1986. p. 91-103.
5. Kleinbaum DG, Kupper L, Morgensten H. Epidemiologic research. Principles and quantitative methods. Nueva York: Van Nostrand Reinhold; 1982.
6. Unidad de Epidemiología Clínica y Bioestadística. Pontificia Universidad Javeriana. Módulos de Bioestadística. Santafé de Bogotá; 1998.
7. Selvin S. Statistical analysis of epiemiologic data. Capítulo 3. Oxford: Oxford University Press; 1991.
8. Converse JM, Presser S. Survey questions. Handcrafting the standardized questionanaire. Series: Quantitative Applications in the Social Sciences. SAGE Publications; 1986.
9. Schoenbach V, Arrighi M. Data management and data analysis. Epid.168. Notas de Clase, Departamento de Epidemiología. Universidad de Carolina del Norte; 1994.
10. Stein JS. Health services research. A pragmatic overview. Capítulos 10-13. 1990. No Publicado.
11. Spilker B. Guide to clinical trials. Filadelfia-Nueva York: Lippincott-Raven; 1996.
12. Bland M. An introduction to medical statistics. Oxford: Oxford Medical Publications; 1995.
13. Wayne WD. Bioestadística. Base para el análisis de las ciencias de la salud. México: Limusa; 1990.
14. Pagano M, Gauvreau K. Principles of biostatistics. California: Duxbury Press; 1993.
15. Kleinbaum DG, Kupper L, Muller K, et al. Applied regression analysis and other multivariable methods. Capítulo 11. Pacific Grove, CA: Duxbury Press; 1998.

Capítulo

30 Modelado

Jhon Camacho

La validez de los resultados y las conclusiones de un estudio clínico dependen no solo de la selección y la ejecución correctas del diseño y de los métodos de análisis, sino, también, de la calidad de los datos. Aun si se emplean técnicas estadísticas avanzadas, es imposible (o, al menos, extremadamente difícil) obtener resultados válidos si la información en la que se basa el análisis no es confiable. Tales riesgos se agudizan en los estudios secundarios, cuyas conclusiones se basan en datos que, en la mayoría de los casos, fueron recopilados o analizados por otros investigadores.

La calidad de los datos es, entonces, fundamental, y garantizarla es una tarea que dista mucho de ser trivial. En particular, debido a que antes de hallarse disponibles para el análisis, los datos suelen haber pasado por varios procesos donde interactúan diferentes actores.

Por fortuna, hay metodologías de manejo de información que previenen estos problemas. Este capítulo y los dos siguientes presentan dichos métodos informáticos, a la vez que proveen ejemplos de su uso para la administración de los datos de un estudio clínico. A lo largo de los tres capítulos, se utiliza como ejemplo un estudio de métodos mixtos que busca entender las barreras que enfrenta una población de pacientes con trasplante renal para adherirse al tratamiento.

Cabe aclarar que los métodos descritos en estos capítulos no constituyen normas prescriptivas, sino, más bien, guías que han comprobado su utilidad para prevenir problemas de calidad de datos, pero aun así admiten ser adaptadas a las condiciones particulares de cada estudio.

La investigación clínica rara vez se hace en solitario. En la mayoría de los casos, hay varias personas involucradas en todos los procesos, desde el diseño del estudio hasta el reporte de los resultados. De hecho, es muy probable que quien planea la pregunta de investigación no sea la misma persona que recolecte los datos o quien analice los datos. Para que ese esfuerzo conjunto funcione, todos los involucrados deben estar de acuerdo acerca de las características de la información del proyecto.

Por ejemplo, una de las variables de interés en nuestro estudio de adherencia[1] era la edad de los participantes. Sin embargo, hay muchas formas de registrar dicha información. La edad

[1] Se trata de un estudio de métodos mixtos, que se usará como ejemplo con el cual se busca entender las barreras que enfrenta una población de pacientes con trasplante renal para adherirse al tratamiento.

podría capturarse en unidades de años cumplidos, o en años y días, o podría registrarse la fecha de nacimiento. Cada opción tiene sus méritos. Para algunos participantes, por ejemplo, los ancianos de áreas rurales, recordar los años cumplidos es mucho más fácil que la propia fecha de nacimiento. Pero esta última provee información más precisa. La opción que se elija va a tener varias implicaciones en el tipo y en la profundidad del análisis. Cuando estas decisiones no son claras ni compartidas por todos los miembros del equipo, es muy común hallar sorpresas desagradables en etapas posteriores del proyecto, y que pueden, incluso, poner en peligro la validez de los resultados.

En este punto, puede ser que algunos de los lectores (particularmente, quienes leen este libro mientras adelantan sus estudios de posgrado) estén pensando: "Bueno…, esto tal vez no aplica para mi estudio... Al fin y al cabo, yo voy a hacerlo casi todo. Yo mismo voy a plantear el diseño, a recoger los datos, y, también, a hacer el análisis". Si este es su caso, lo invito a que lo piense de nuevo. Muchas cosas van a pasar desde la definición de la pregunta hasta la escritura de los resultados.



Construir y documentar esas definiciones desde el inicio del proyecto ayuda a que las decisiones que se tomen por el camino sean coherentes con el resultado que se busca, y previene problemas que son muy difíciles de solucionar cuando ya ha pasado la fase de recolección.

Una forma de construir y documentar las definiciones acerca de los datos y documentarlas para que puedan ser compartidas y mantenidas a lo largo del estudio es la construcción de un modelo de los datos.

MODELADO DE LOS DATOS

El modelo de datos responde preguntas como:

- ¿Qué datos se van a recoger y a almacenar?
- ¿Cómo se relacionan esos datos entre sí?
- ¿Qué reglas deben cumplir los datos para considerarse válidos?

Por ejemplo, un modelo de datos para el estudio de adherencia debería definir, entre otros aspectos, si se va a capturar la información de las instituciones que han tratado al participante desde su trasplante, o si, por el contrario, solo se va a capturar la información de la institución que lo trata actualmente. También debería definir si todo paciente debe tener asociada la información de la institución tratante o si pueden aceptarse pacientes sin esa información. Así mismo, debería definir cómo identificar cada institución de forma que no puedan confundirse, y que la misma institución siempre se identifique de la misma forma.

Este último punto merece un poco más de detalle. A simple vista, parecería que la forma más simple de identificar la institución tratante es a través del nombre, preguntado a cada participante. El problema con eso es la posibilidad de que los pacientes tratados en el mismo centro utilicen nombres diferentes (al menos, ligeramente) para referirse a este. Uno de ellos podría hablar del "Hospital Universitario San Ignacio", mientras que otro podría hablar del "HUSI", o de la Clínica San Ignacio. En ese caso, para un ojo no entrenado, estos podrían ser hospitales diferentes. Una mejor opción es construir una lista predefinida de instituciones, identificadas por un código numérico único, y pedirle al participante que seleccione la suya propia de la lista. Aun así, es posible que algunos participantes no reconozcan su hospital en la lista, por lo cual podría ser necesario incluir la opción "Otra institución" y una variable de texto abierto para ingresar el nombre del hospital.

Componentes del modelo de datos



El modelo de datos define la estructura de la base de datos del estudio y las reglas que deben cumplir los valores de las variables.

Hay varias metodologías para construir y documentar estas definiciones. Sin embargo, casi todas coinciden en que, como mínimo, el modelo de datos debe contener tres tipos de componentes: 1) las entidades a las que pertenecen los datos, 2) los atributos de las entidades que son de interés para el modelo y 3) las relaciones que hay entre sí en las entidades. Las siguientes

secciones exploran cada uno de dichos componentes, pero antes de continuar, hablemos un poco acerca de la documentación.

EJEMPLO

Al lector interesado en profundizar en los componentes del modelo de datos se le recomienda *Data Modeling Made Simple: A Practical Guide for Business and IT Professionals*, de Steven Hoberman[1]. A pesar de que, por el título, parecerá un libro para ingenieros, Hoberman logra explicar de manera simple y práctica estos conceptos, a través de ejemplos que no requieren conocimientos previos de ingeniería.

Las entidades, los atributos y las relaciones del modelo se suelen representar con un diagrama como el que se muestra en la **figura 30.1**. Diagramas como este reciben el nombre de *diagramas de entidad-relación* (ERD, por las siglas en inglés de ***entity-relationship diagram***). Hay una buena cantidad de símbolos en ese gráfico, y puede que en este punto no sea claro el significado de varios de ellos. Pero no hay por qué desesperar: a medida que se hable de cada componente del modelo, su significado será más claro. Por ahora, basta con decir que el mencionado diagrama dice que la base de datos del estudio almacenará información acerca de las mediciones en sangre de ciertos medicamentos, y que cada medición pertenece a un paciente y tiene asociada una fecha y un valor, llamado ***nivel en sangre***.

EJEMPLO

El diagrama de la **figura 30.1** se basa en un estándar llamado *notación Barker*, propuesto durante la década de los 1980 por Richard Barker *et al*. Al lector interesado se le recomienda revisar el libro *case Method: Entity Relationship Modelling*[2]. Hay muchas aplicaciones que facilitan la creación de estos diagramas. Los que se usan en el presente capítulo fueron construidos utilizando las herramientas del sitio web http://creately.com

Entidades

El primer paso en la construcción del modelo de datos consiste en identificar los conceptos (personas, animales, cosas, sucesos) que son de interés para el estudio y sobre los cuales se almacenará información en la base de datos. Tales conceptos reciben el nombre de ***entidades***.

Las entidades representan colecciones de objetos, sucesos o actores que interactúan u ocurren para conformar el proceso acerca del cual se va a capturar información.

Por ejemplo, el diagrama de la **figura 30.1** representa la información acerca de las mediciones de los niveles en la sangre de algunos medicamentos de interés para el estudio de adherencia, y contiene tres entidades: 1) Participantes, 2) Mediciones y 3) Medicamentos; cada una de ellas, representada por una caja.

Un método útil para identificar las entidades que debe contener el modelo consiste en preguntarse qué datos se requieren para responder la pregunta de investigación, y luego, preguntarse a qué tipo de cosa o de persona pertenecen esos datos. La respuesta a esta última pregunta es la entidad. Por ejemplo, para llevar un control de la recolección de datos cualitativos del estudio de adherencia, puede ser necesario almacenar la fecha en la que se entrevistó a cada participante, el nombre del entrevistador, el nombre del transcriptor y la localización del archivo con la transcripción. Esos datos podrían representarse a través de las entidades: participantes, entrevistas, entrevistadores y transcriptores. Es posible que en este punto no sea claro si se debe o no incluir una entidad llamada ***archivo de trascripción***, o si los participantes, los entrevistadores y los transcriptores deberían reunirse en una sola entidad llamada ***personas***.

Sobre el modelado, hay dos cosas importantes que decir. Primero, el proceso de modelado es ***iterativo***; es muy común que durante el siguiente paso (durante el cual se definen los atributos de las entidades) se agreguen o se eliminen entidades; y segundo, el modelado no es una ciencia exacta; dos personas pueden definir modelos perfectamente razonables (aunque diametralmente diferentes) para los mismos datos. Debido a ello, no hay recetas estrictas para construir el modelo, sino solo algunas guías útiles. Habiendo dicho esto, también cabe reconocer que hay modelos más útiles y fáciles

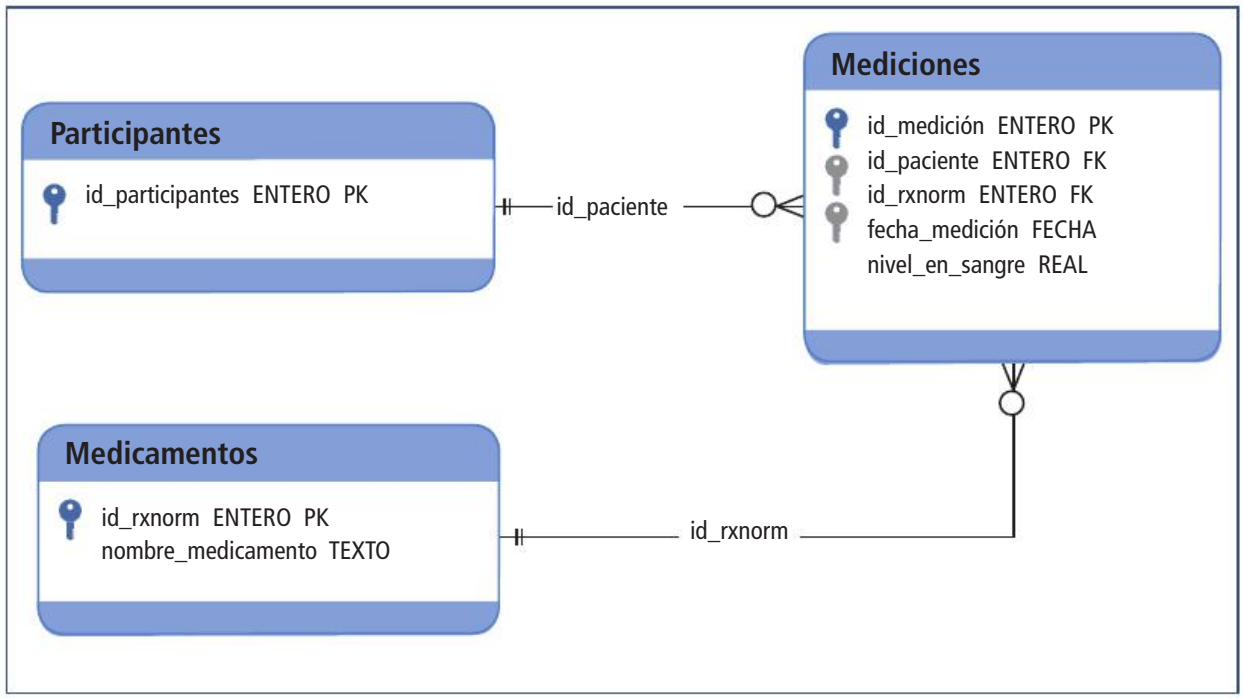


Figura 30.1 ERD correspondiente a las mediciones de los niveles en sangre.

de usar que otros. Al lector interesado en la evaluación de modelos de información médica se le recomienda revisar ***Data Model Considerations for Clinical Effectiveness Researchers***, de Kahn ***et al.***[3].

EJEMPLO

¿Será esta una entidad?

Un error común al iniciarse en el mundo del modelado de datos es confundir las entidades con sus instancias. Mientras una *entidad* se refiere a una colección, las *instancias* son los elementos que constituyen la colección. Si la entidad fuera una tabla en una hoja de cálculo, entonces las instancias serían las filas de la tabla.

Otro problema usual es confundir las entidades con los procesos. Mientras los *procesos* se refieren a un conjunto de actividades que suceden en el tiempo, las *entidades* son los componentes que interactúan para que suceda el proceso. Por ejemplo, la información del proceso de atención en servicio de urgencias de un hospital podría ser modelado a través de entidades que podrían incluir médicos, enfermeras, pacientes, diagnósticos y procedimientos. Esto, sin embargo, no impide que otro modelo alusivo a los diferentes procesos de un hospital contenga una entidad llamada *proceso*, y de la cual *urgencias* sea una *instancia*.

Atributos

Los *atributos* son las características de las entidades que ayudan a resolver la pregunta de investigación. En otras palabras, los atributos son las variables de las cuales se va a capturar y a almacenar información, y cuyos valores van a ser la materia prima del análisis.

Los atributos definen el contenido de la base de datos, y sus valores describen las instancias de las entidades. La **figura 30.2** muestra la entidad ***participantes***, que incluye algunos de los atributos de interés para el estudio, iniciando por ***id_participante***, cuyo valor es un número entero, y que servirá para identificar a cada sujeto (más adelante se hablará del símbolo de la llave y de la sigla **PK**; por ahora, dejémoslas pasar). Adicionalmente, la entidad Participantes va a incluir información acerca del nombre, el sexo, la fecha de nacimiento y el número de años en diálisis.

Hablemos un poco acerca los nombres de los atributos. En términos generales, lo más útil es seleccionar una palabra o una frase que muestre con claridad qué información se almacenará. A veces, esto es más fácil de decir que

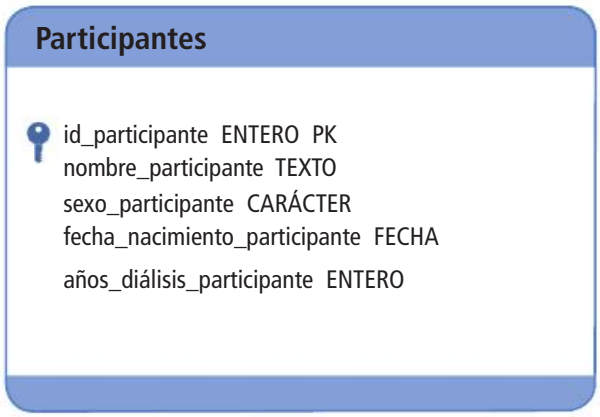


Figura 30.2 **Entidad** *Participantes*, incluyendo algunos de sus atributos.

de hacer. La recomendación en tal sentido es discutir el modelo con todos los actores interesados y cambiar los nombres cuantas veces sea necesario, hasta cuando todos entiendan con facilidad su significado. Habiendo dicho esto, según la propia experiencia del autor, la mayoría de las veces no son necesarios muchos cambios. Lo importante es que para todos sea fácil entender lo que busca expresar el diagrama.

Los nombres de atributo de la **figura 30.2** usan algunas convenciones que facilitan la lectura del diagrama:

- Los espacios en blanco se reemplazan por "_" para ayudar a diferenciar el nombre de otras características del atributo también incluidas en el diagrama (como el tipo de dato, escrito en mayúsculas; un poco más adelante, hablaremos de los tipos de dato).
- El nombre del atributo se finaliza con el nombre de la entidad en singular. Esto logra dos cosas: 1) se elimina la ambigüedad que ocurre cuando dos entidades necesitan el mismo nombre de campo (el ejemplo típico es el campo id, el cual suele usarse para identificar cada instancia); y 2) es suficiente con referirse al nombre del campo para saber a qué entidad pertenece, lo que va a ser muy útil al definir las relaciones entre entidades.

Además del nombre de cada atributo, el modelo de datos también necesita definir sus valores permitidos. Parte de dicha información está representada en el ERD, a través de las definiciones de tipo de dato que aparecen escritas en mayúsculas a continuación del nombre. Por ejemplo, el diagrama de la **figura 30.2** estipula que el atributo id_participante solo admite números enteros, mientras que sexo_participante admite un carácter. A igual que con los nombres de los atributos, el nombre del tipo de dato no necesita seguir una regla estricta: lo importante es que todos los interesados entiendan su significado.

Volviendo al atributo sexo_participante, lo que no dice el diagrama es qué carácter representa el sexo masculino o si es posible ingresar datos sobre un participante con ambigüedad sexual. Una forma de documentar esta información es a través de una tabla de definición de valores, como la **tabla 30.1**.

Una tabla de definición de valores, como la que se muestra en la **tabla 30.1**, complementa el ERD con información acerca de los valores permitidos para el atributo, su significado, su nivel de requerimiento y su formato.

Tabla 30.1 **Definición de valores para los atributos de la entidad Participantes**

Atributo	Tipo de dato	Valores	Formato	Significado	Obligatorio
id_participante	Entero	> 0			Sí
nombre_participante	Texto	Máximo, 30 caracteres			Sí
sexo_participante	Carácter	M, F		*M* = Masculino *F* = Femenino	Sí
fecha_nacimiento_participante	Fecha	≥ 01/ENE/1900, y ≤ 01/ENE/2020	DD/MMM/AAAA		Sí
años_diálisis_participante	Entero	> 0 y < 21			No

Fuente: Elaboración propia.

Por ejemplo, estipula que el atributo id_participante es obligatorio; o sea, no admite valores nulos o quedar vacío, y solo puede contener números enteros > 0. Por el contrario, el atributo años_diálisis_participante puede quedar vacío, pero solo admite valores enteros > 0 y < 21. Acerca de sexo_participante, la tabla estipula que admite solo dos valores: M o F; el primero, para el sexo masculino, y el segundo, para el sexo femenino. Finalmente, algunos atributos, como los de tipo FECHA, pueden requerir información de formato.

EJEMPLO

¿Cálculos o datos para cálculos?

A algunos lectores les puede llamar la atención que la entidad Participantes incluya la fecha de nacimiento en lugar de la edad. Una razón para esto es que a partir de la fecha de nacimiento se puede calcular la edad con diferentes grados de precisión (por ejemplo, en años o en años y meses). Adicionalmente, la edad es un dato cambiante que depende de la fecha contra la cual se compare la fecha de nacimiento. Así pues, el dato de edad por sí mismo no es suficiente: habría que decir algo como "edad del participante en la fecha X", y esto, claramente, complica las cosas. Cuando se trata de capturar información, en general, es más útil capturar los datos que dan origen a los cálculos que su resultado. En otras palabras, normalmente, es mejor capturar los datos en su estado más puro.

Identificadores

Todo estudio necesita una forma de identificar a cada participante, y es posible que ni el nombre ni el documento de identidad estatal sean los medios más apropiados para ello. Identificar a los participantes por el nombre puede crear problemas con homónimos, e identificarlos a través del documento de identidad estatal puede afectar la confidencialidad de su información. Tal necesidad de identificar unívocamente cada instancia de una entidad no solo se restringe a las personas. Cada instancia necesita que se la pueda identificar dentro de la entidad, para poder conectar y analizar los datos. Por ejemplo, en el ERD acerca de las mediciones de niveles en sangre (repetido en la **figura 30.3**), es necesario identificar cada medicamento y cada medición.

Esta tarea de diferenciar las instancias de las entidades se logra a través de los ***atributos identificadores***.

Un identificador es un atributo cuyos valores no se repiten entre las instancias de una entidad, y gracias a eso, sirve para diferenciar cada una de ellas sin peligro de ambigüedad.

Cualquier atributo puede servir de identificador, siempre y cuando sus valores sean únicos. Incluso, se pueden usar combinaciones de atributos como identificador. Sin embargo, por convención, se suele usar un atributo cuyo nombre inicia con la sigla ID (*identificación*, de uso internacional) y almacena un número entero. La sigla busca que sea fácil ubicar el identificador de cada entidad, lo que, de nuevo, será muy útil a la hora de definir relaciones. En cuanto al

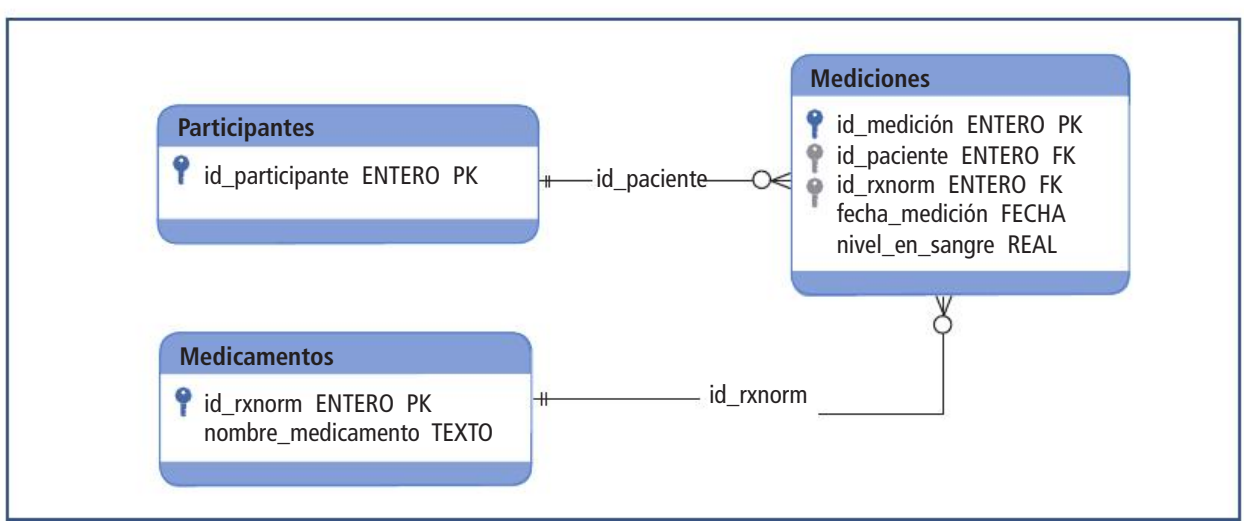


Figura 30.3 ERD correspondiente a las mediciones de los niveles en la sangre.

número entero, se debe a que tarde o temprano el modelo se va a convertir en una serie de tablas en una base de datos, y, en general, para los computadores es mucho más fácil (y, por ende, más rápido) comparar dos números enteros que comparar cualquier otro tipo de dato.

EJEMPLO

¿Llaves?
Otra forma de nombrar los atributos identificadores es llamarlos *Llave primaria*. El nombre hace referencia a que estos atributos sirven para acceder a una instancia de entidad determinada. En los ERD, se suele representar los identificadores con un símbolo con forma de llave o con la sigla PK (*primary key*).

Relaciones

Los datos cualitativos de nuestro estudio de adherencia se obtuvieron a través de entrevistas individuales en las que se exploraron las barreras que enfrentaban los participantes para seguir las recomendaciones del personal de salud. Cada entrevista incluyó solo a un participante, y fue conducida por un solo investigador. Las entrevistas fueron grabadas y luego transcritas. Además del participante, del nombre del investigador y de la fecha de la entrevista, también se necesitaba llevar un registro de los nombres de los archivos de las transcripciones, para facilitar la organización del proyecto de análisis cualitativo. El modelo de entidad-relación de la **figura 30.4** representa estas definiciones.

La mayoría de los componentes del diagrama de la **figura 30.4** ya deben ser familiares para el lector; sin embargo, hasta ahora no hemos hablado del significado de las líneas que conectan las entidades. Estos símbolos representan las relaciones entre las entidades del diagrama.



Las relaciones definen las reglas que deben cumplir las instancias de las entidades al relacionarse entre sí.

Por ejemplo, el diagrama de la **figura 30.4** estipula que cada entrevista debe estar relacionada con uno y solo con un participante, y que cada participante puede estar relacionado con cero o más entrevistas. Tal definición acerca del número de instancias que participan en la relación se llama ***cardinalidad***. Por convención, las relaciones se nombran de acuerdo con su cardinalidad. Por ejemplo, entre Participantes y Entrevistas existe una relación "de uno a muchos". Alguien podría decir que también podría llamarse de "muchos a uno", pero, de nuevo por convención, se suele decir que es de "uno a muchos". La **tabla 30.2** presenta los símbolos de cardinalidad más comúnmente usados.

Además de los símbolos de cardinalidad, la relación también incluye un texto que hace referencia al identificador que conecta las dos entidades. Por ejemplo, la **figura 30.4** muestra que la relación entre Entrevistas e Investigadores se mantiene a través de atributo id_investigador que identifica cada instancia de Investigadores, pero también se encuentra en la entidad Entrevistas. Incluir este identificador en la entidad Entrevistas permite encontrar al investigador relacionado con una entrevista dada.

EJEMPLO

¿Llaves foráneas?
Un identificador que se incluye en otra entidad para generar una relación toma el nombre de *Llave foránea*, y suele representarse con la sigla FK (*foreign key*).

Un par de patrones útiles

Entidades, atributos y relaciones son todo lo que se necesita para construir un modelo de datos para la mayoría de los estudios clínicos. La combinación exacta de estos elementos depende no solo de la creatividad de quien construye el modelo, sino también, de las necesidades particulares del estudio; y si bien cada estudio viene con sus propios requerimientos de información, se ha encontrado que algunos problemas son generales. La solución a dichos problemas ha dado origen a combinaciones específicas de elementos, que han probado ser de utilidad en muchos escenarios. Estas combinaciones reciben el nombre de ***patrones de diseño***. Una descripción detallada de los patrones de diseño en el área de modelado de datos (incluso una enumeración de todas ellas) escapa del alcance del presente capítulo. Sin embargo, a continuación se describen dos patrones que, en opinión del autor, todo modelador de datos debe conocer, y que ayudan a prevenir los problemas

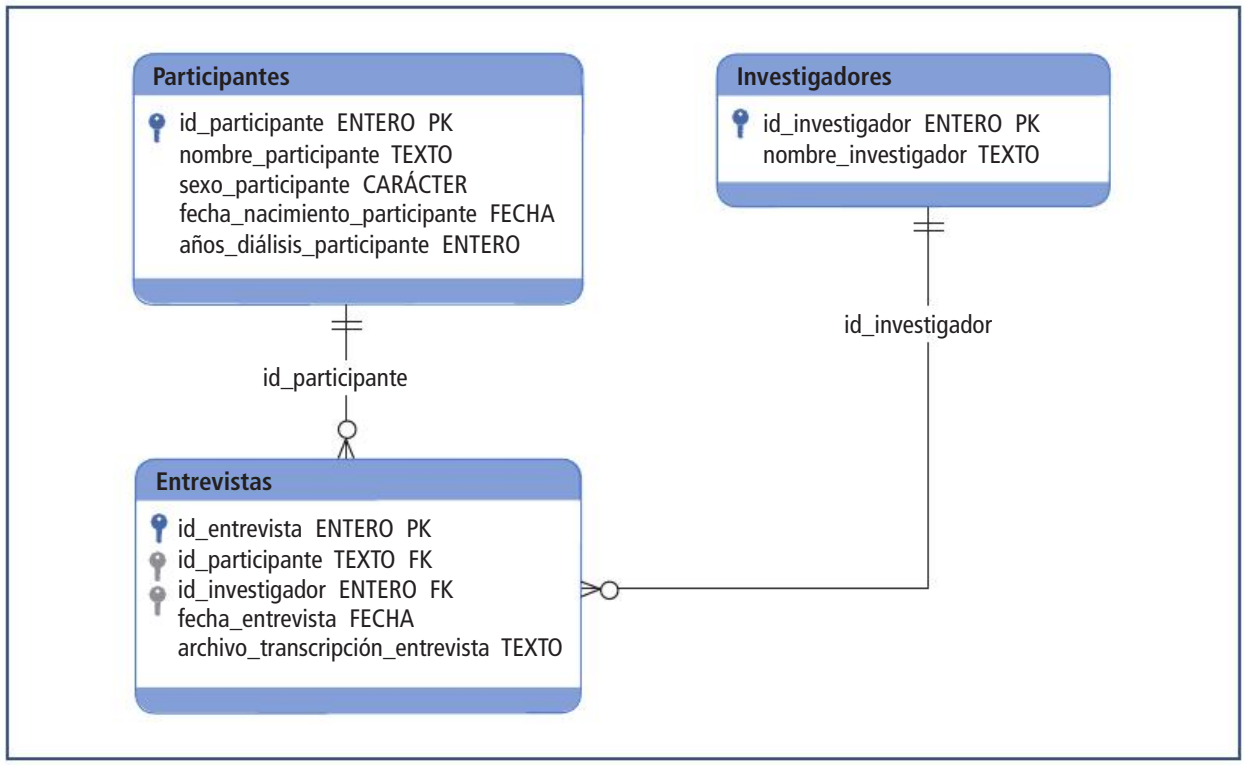


Figura 30.4 ERD correspondiente a la información de control de las entrevistas cualitativas.

más frecuentes en las bases de datos de estudios clínicos.

EJEMPLO

Al lector interesado en profundizar acerca de patrones de diseño para modelos de datos se le recomienda revisar *Relational Database Design Clearly Explained*, de Jan L. Harrington[4], y *Data Model Patterns: A Metadata Map*, de David C. Hay[5].

Relaciones de muchos a muchos

La técnica de incluir el identificador de una entidad en otra, para definir una relación, funciona en la mayoría de los casos; sin embargo, hay un caso que requiere atención especial: las relaciones de muchos a muchos. El problema con dichas relaciones es que cada instancia involucrada necesita hacer referencia a más de una instancia de otra entidad, por lo cual un solo atributo no es suficiente.

Tabla 30.2 Símbolos de cardinalidad más comúnmente usados

Símbolo	Significado
	Uno y solo uno
	Cero o uno
	Uno o muchos
	Cero o muchos

Fuente: Elaboración propia.

La solución más sencilla para modelar una relación de muchos a muchos es incluir una tercera entidad que medie la relación, como se ve en la **figura 30.5**.

El diagrama de la **figura 30.5** representa la relación que hay entre los participantes de nuestro estudio de adherencia y los profesionales de la salud responsables de su tratamiento. El problema con esta relación es que un mismo participante puede ser atendido por más de un profesional de la salud, y simultáneamente, un mismo profesional de la salud puede atender a más de un participante. En otras palabras, se trata de una relación de muchos a muchos. La solución que se expone en la **figura 30.5** consiste en incluir una entidad llamada Relaciones_Participantes_Profesio-

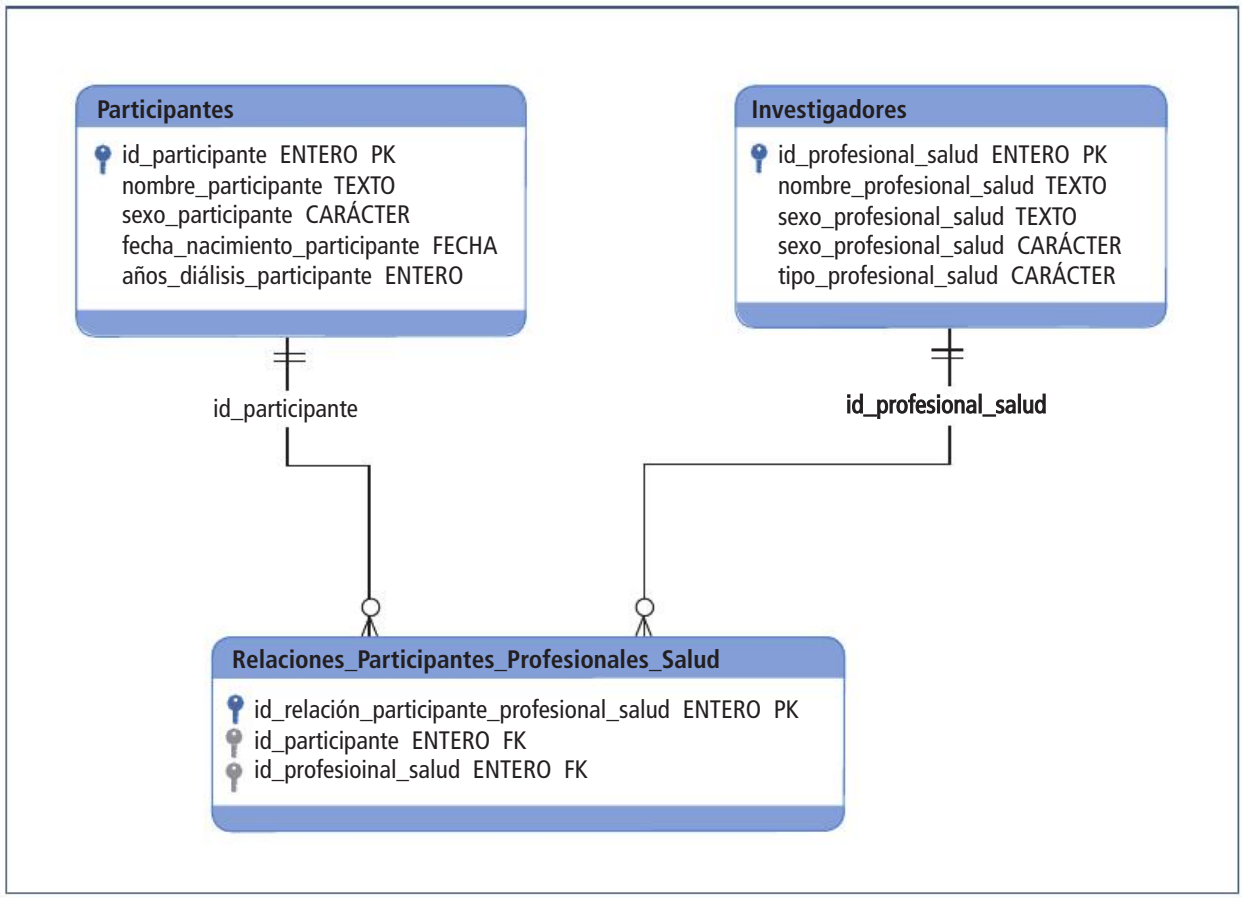


Figura 30.5 Ejemplo de entidad mediadora para resolver relaciones de muchos a muchos.

nales_Salud, de forma que cada instancia de la nueva entidad represente la relación entre un único participante y un único profesional de la salud.

Valores de referencia

Al inicio del capítulo, hablábamos de los retos relacionados con la captura de la información acerca de la institución tratante en nuestro estudio de adherencia. El problema era que los participantes podrían referirse a la misma institución con nombres diferentes. Dos participantes tratados en el Hospital Universitario San Ignacio podrían referirse a dicha entidad como el "HUSI" o como "San Ignacio". Si tal situación no se corrige antes del análisis, casi cualquier cálculo estadístico va a considerar ambos nombres pertenecientes a dos hospitales diferentes, lo que, claramente, afectaría la validez de los resultados. Imagine ahora el lector un estudio donde no se hace referencia a un hospital, sino a 30, o donde la información de la institución se captura como una pregunta abierta en texto, lo cual permite nombrar el mismo hospital como "San Ignacio" o "S. Ignacio" u "H. S. Ignacio". Es típico en estos escenarios que un conteo de frecuencias encuentre no 30, sino 300 o 500 instituciones. Imagine el lector cuánto trabajo requeriría arreglar este problema si, además, pasa lo mismo con varios atributos.

Afortunadamente, hay una solución relativamente sencilla para el problema descrito (siempre y cuando se lo implemente antes de la fase de captura de los datos). La técnica consiste en incluir una entidad que alberge los valores de referencia del atributo, como se ve en la **figura 30.6.** El modelo de la **figura 30.6** ayuda a resolver el problema de la captura del nombre de la institución tratante estipulando una única forma de referirse a cada institución.



La inclusión de una entidad que agrupe los valores de referencia de un atributo ayuda a evitar la ambigüedad entre los posibles valores.

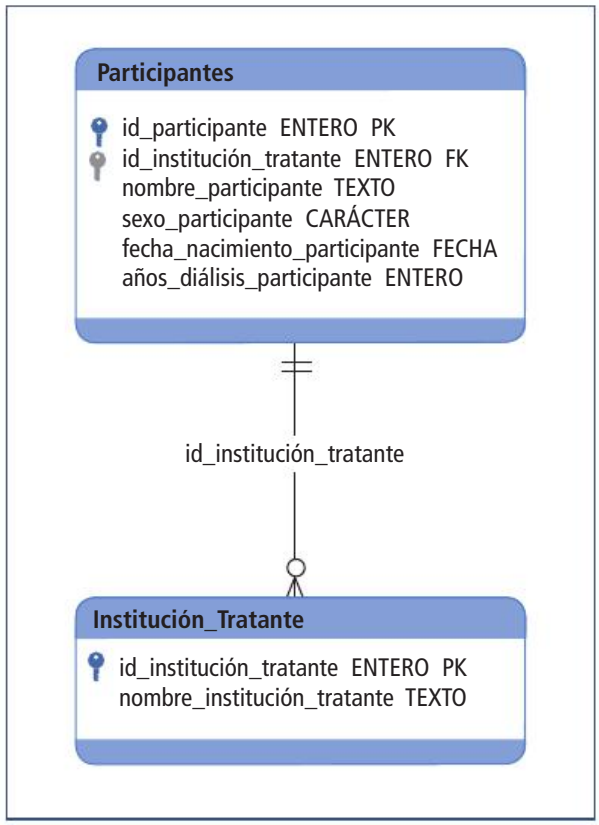


Figura 30.6 Ejemplo de la inclusión de valores de referencia ERD acerca de la información de la institución tratante.

Terminologías médicas

La técnica de los valores de referencia se puede llevar un paso más allá mediante el uso de terminologías médicas. Por ejemplo, en nuestro estudio sobre adherencia nos interesaba registrar los medicamentos que debía tomar cada participante. El problema es que al inicio del estudio no teníamos una lista de todos los posibles medicamentos que podría tomar un paciente que ha pasado por un trasplante. La solución provino de usar la terminología médica RxNorm[6], como se ve en la **figura 30.7**.

Una terminología médica es un estándar que define términos y reglas de asociación acerca de un cuerpo de conocimiento médico.

Por ejemplo, RxNorm contiene una lista estandarizada de nombres de medicamentos, incluyendo sus nombres comerciales en Estados Unidos. Incluir RxNorm en nuestro estudio requirió descargar los archivos desde el sitio web de la US National Library of Medicine (NLM) y cargarlos a nuestra base de datos (se tratará más sobre esto en el **capítulo 31** "Construcción de bases de datos", de este mismo libro). La **tabla 30.3** lista algunas otras terminologías médicas, junto con su área de conocimiento, la entidad responsable y el sitio web donde se puede obtener más información.

Terminologías como las de la **tabla 30.3** son el resultado de la colaboración de muchos entes comerciales, gubernamentales y académicos, y son constantemente actualizadas. Adicionalmente, muchos sistemas de información médica utilizan esas terminologías para facilitar el intercambio de información entre ellos. Esto significa que al incluirlas en el modelo de datos no solo se logran los beneficios de los valores de referencia, sino que también se facilita el uso secundario de los datos del estudio.

EJEMPLO

Al lector interesado en profundizar en el tema de terminologías médicas se le recomienda revisar *Estándares en informática médica: generalidades y aplicaciones*, de Suárez-Obando y Camacho[11].

MODELADO DE LA CAPTURA DE LOS DATOS

Hasta ahora, hemos visto cómo modelar el estado en el que deben quedar los datos luego de ser capturados. Pero, ¿qué pasa con el proceso de captura? Esta sección se enfoca en cómo plasmar en un modelo de captura de datos las decisiones (muchas veces, tácitas) que se toman durante la construcción de los cuestionarios.

Un modelo de captura de datos especifica la secuencia de acciones que se debe llevar a cabo para capturar los datos, a partir de las definiciones hechas en los cuestionarios.

La **tabla 30.4** presenta una porción del cuestionario de caracterización de sujetos empleado en nuestro estudio de adherencia. Este ejemplo solo tiene cinco preguntas, y, a pesar de ello, incluye una buena cantidad de definiciones que no conviene pasar por alto. Dice, por ejemplo, que solo se debe capturar información de sujetos cuyas edades estén en el rango de los 15 a los

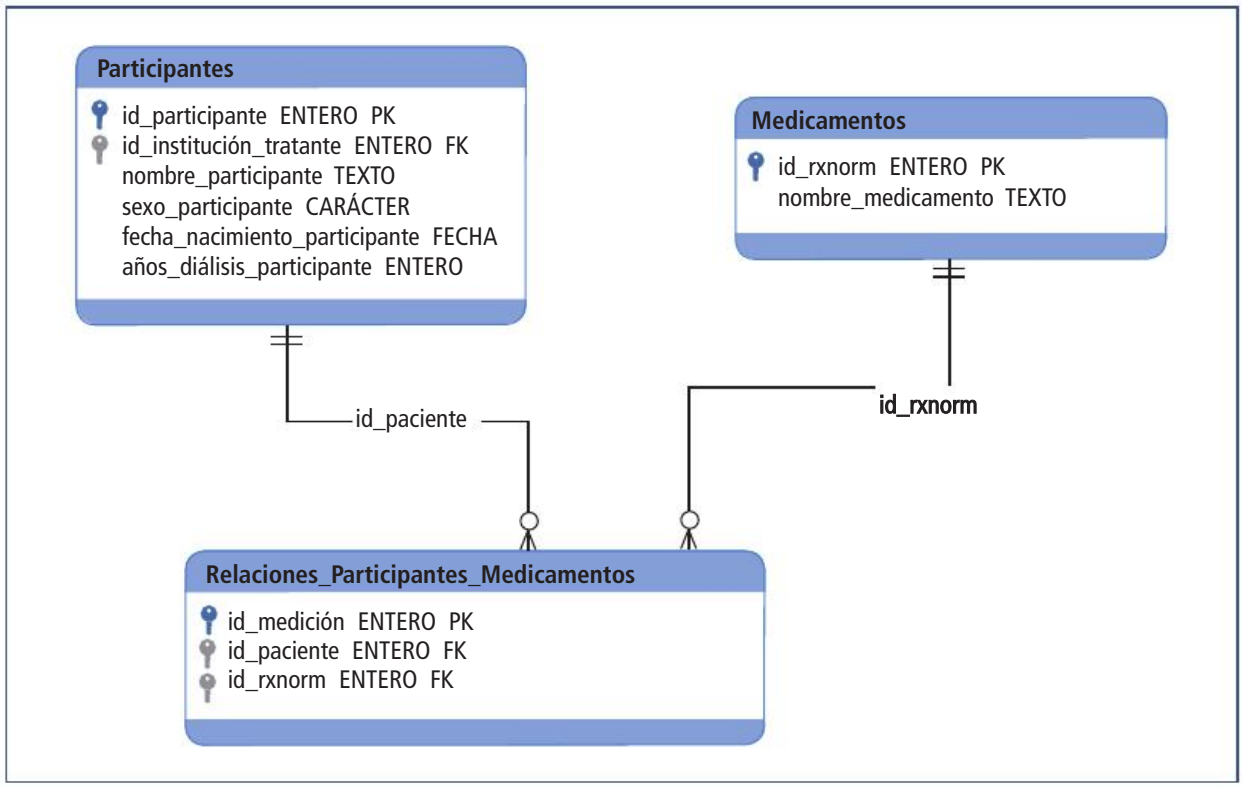


Figura 30.7 **Inclusión de una terminología médica. Diagrama de entidad-relación correspondiente a los medicamentos que toma** cada participante.

Tabla 30.3 Ejemplos de terminologías médicas

Área de conocimiento	Terminología	Entidad responsable	Dirección web
Diagnóstico	**International Classification** of Diseases (IDC)[7]	Organización Mundial de la Salud (oms)	**http://www.who.int/** classifications/icd/en/
Diagnóstico	**Systematized Nomenclature** of Medicine-Clinical Terms (SNOMED-CT)[8]	International Health Terminology Standards Development Organization (IHTSDO)	http://www.ihtsdo.org/snomed-ct
Tópicos de literatura médica	**Medical Subject Headings** (MeSH)[9]	us **National Library of** Medicine (NLM)	**http://www.nlm.nih.gov/mesh/** meshhome.html
Medicamentos	**RxNorm**[6]	us **National Library of** Medicine (NLM)	**http://www.nlm.nih.gov/research/** umls/rxnorm/
Observaciones (procedimientos diagnósticos)	**Logical Observation** Identifiers Names and Codes (LOINC)[10]	Regenstrief Institute	http://loinc.org/

Fuente: Elaboración propia.

100 años, y que se debe capturar la información de medicamentos con base en una lista y una pregunta de respuesta múltiple.

Componentes del modelo de captura

En el cuestionario del ejemplo se pueden identificar tres tipos de elementos: mensajes, preguntas y saltos condicionales. A través de la identificación y la documentación de dichos elementos, se pueden integrar las características del formulario en un modelo de captura que facilite la discusión y la comprensión de estas decisiones entre los miembros del equipo del estudio.

Una forma útil de representar el modelo de captura es mediante el uso de un diagrama de flujo,

Tabla 30.4 Ejemplo de cuestionario

Caracterización de sujetos-cuestionario para participantes									
Las siguientes preguntas nos ayudarán a saber un poco acerca de su experiencia como paciente con trasplante renal, con el fin de ajustar las preguntas de la entrevista a la que le hemos invitado									
1	¿Cuál es su fecha de nacimiento?			Día		Mes		Año	
Pase a validación 1									
Validación 1	**¿La edad del sujeto es igual o** mayor a 15 años y es menor de 100 años?	1		Sí		*Pase a 2*			
		2		No		*El sujeto no es elegible (Terminar)*			
2	¿Cómo calificaría la atención médica que ha recibido desde cuando fue trasplantado?	1	Excelente			*Pase a 4*			
		2	Buena			*Pase a 4*			
		3	*Regular*			*Pase a 3*			
		4	*Mala*			*Pase a 3*			
		5	*Muy Mala*			*Pase a 3*			
		99	*No responde*			*Pase a 3*			
3	¿Cuál es el principal factor que le hace sentir que no es buena la atención médica que ha recibido desde su trasplante?								
	Pase a 4								
4	¿Qué medicamentos toma diariamente? (marque todas las opciones que apliquen)	**1**	**Ácido micofenólico**						
		2	Gemfibrozilo®						
		3	Tacrolimus						
		4	Ciclosporina						
		5	Sirolimus						
		98	*Otro*			*Pase a validación 2*			
Validación 2	¿La respuesta a la pregunta 4 incluye "Otro"?	**1**	**Sí**			*Pase a 5*			
		2	No			*Terminar*			
5	¿Qué otros medicamentos toma diariamente?								
						Terminar			

Fuente: Elaboración propia.

como el que se presenta en la **figura 30.8**. Al comparar el diagrama de la figura con el cuestionario de la **tabla 30.4**, debe ser evidente para el lector que el diagrama representa el inicio del formulario hasta el punto en el cual se hace la "Validación 1", que se encarga de rechazar a los sujetos que no estén dentro del rango de edad especificado.

El diagrama de la **figura 30.8** contiene cinco elementos, que representan las entradas y las salidas del diagrama, así como los mensajes, las preguntas y los saltos condicionales correspondientes a esta sección del cuestionario. El significado de cada símbolo se presenta en la **tabla 30.5**. Las siguientes secciones exploran cada uno de los mencionados componentes.

EJEMPLO

Al igual que con los ERD, los diagramas de flujo presentados en esta sección fueron construidos utilizando las herramientas del sitio web http://creately.com

El diagrama de flujo facilita la discusión acerca de la secuencia de pasos que llevan a la captura de los datos; principalmente, porque

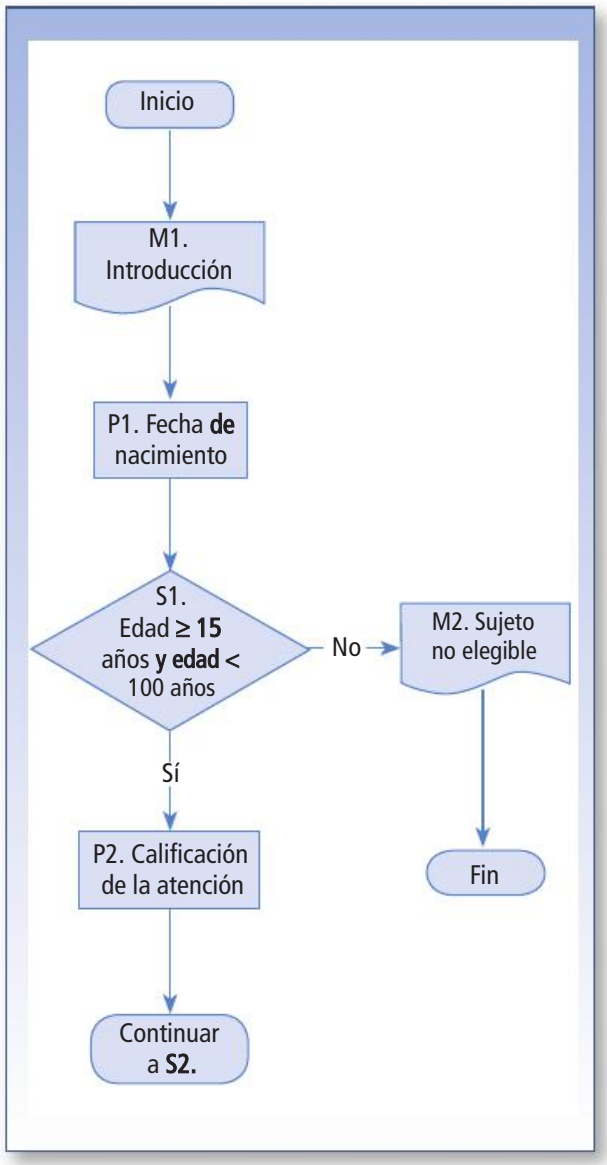


Figura 30.8 Diagrama de flujo correspondiente al inicio del cuestionario y la validación de la edad del sujeto.

permite visualizarla en conjunto. Es fácil ver en el diagrama de la **figura 30.8** que existen dos caminos de acción, y que la decisión acerca de cuál camino seguir depende de la edad del sujeto. Sin embargo, es posible que el lector haya notado que hay información del cuestionario que no aparece en el diagrama, como, por ejemplo, el enunciado del mensaje de introducción. Debido a ello, es útil acompañar el diagrama de flujo con una tabla de definición de pasos, como la que se presenta a continuación.

La **tabla 30.6** contiene toda la información del diagrama de la **figura 30.8** y del inicio del cuestionario de la **tabla 30.4**. Cada uno de estos documentos tiene sus méritos: el cuestionario es fácil de entender y de diligenciar para una persona que no esté involucrada en el manejo de los datos del estudio; el diagrama de flujo permite visualizar claramente la secuencia de pasos; y la tabla de definición de pasos descompone los elementos del cuestionario y los detalla, lo que facilita controlar la ejecución de estos pasos durante la fase de captura.

Tabla 30.5 Símbolos de flujo más comúnmente usados en el modelamiento de captura de información

Símbolo	Significado
	Entrada o salida
	Mensaje
	Pregunta
	Salto condicional

Fuente: Elaboración propia.

EJEMPLO

¿Adónde se fueron las opciones de respuesta?

Es posible que al observar la **tabla 30.6**, algunos lectores echen de menos las opciones de respuesta de la pregunta P2. Lo que sucede es que, en vez de incluir las opciones de respuesta, la tabla de definición de pasos hace referencia al atributo en el que se almacenará la información de la pregunta, cuyos valores se establecen en la tabla de definición de valores. De esta forma, la información acerca de los posibles valores (y, por ende, la de las opciones de respuesta) solo existe en una parte del modelo. Repetir dicha información en los dos modelos (el de datos y el de captura) dejaría abierta la puerta para que las opciones de respuesta no coincidieran con los valores permitidos del atributo.

Más patrones útiles

Al igual que en el modelado de datos, en la captura también aparecen situaciones que se repiten una y otra vez a lo largo de diferentes estudios, y para las cuales hay soluciones prácticas que constituyen patrones de diseño. A continuación, dos de las más comunes.

Preguntas de profundización

Las preguntas 2 y 3 del ejemplo (reproducidas en la **tabla 30.7**) exploran la percepción sobre

Tabla 30.6 Tabla de definición de los pasos correspondientes al inicio del cuestionario y la validación de la edad del sujeto

Tipo	ID	Enunciado	Atributo	Valor	Sig. Paso
Mensaje	M1	**Las siguientes preguntas nos** ayudarán a saber un poco acerca de su experiencia como paciente con trasplante renal, con el fin de ajustar las preguntas de la entrevista a la que lo hemos invitado			P1
Pregunta	P1	**¿Cuál es su fecha de nacimiento?**	*fecha_nacimiento_participante*		S1
Salto Condicional	S1	¿La edad del sujeto es igual o mayor a 15 años, es menor de 100 años?		*1 = Sí*	P2
				2 = No	M2
Mensaje	M2	**El sujeto no es elegible**			**FIN**
Pregunta	P2	**¿Cómo calificaría la atención médica** que ha recibido desde cuando fue trasplantado?	*calificación_atención_participante*		S2

Fuente: Elaboración propia.

la calidad de la atención médica recibida por el participante. La pregunta hace una exploración inicial, mientras que la pregunta 3 profundiza en los factores asociados a la percepción del sujeto, en caso de que este califique la calidad de la atención por debajo del nivel "Buena".

Este tipo de pregunta suele modelarse a través de la combinación de elementos que se muestra en el diagrama de la **figura 30.9**.

El punto clave de este patrón consiste en ubicar la pregunta de profundidad inmediatamente después de la pregunta de exploración. De esa forma, solo es necesario "saltar" la pregunta de profundidad en el caso de las opciones de respuesta "triviales" de las pregunta de exploración.

Dos puntos adicionales para tener en cuenta:

- Si bien la pregunta de profundidad del ejemplo es abierta, ello no es obligatorio. El patrón funciona igualmente bien con cualquier otro tipo de pregunta de profundización.
- Así mismo, si bien el ejemplo solo incluye una pregunta de profundización, nada en el patrón impide incluir más de una pregunta. El cambio se limita a ampliar el salto.

Preguntas con respuesta múltiple

El enunciado de la preguntas 4 del ejemplo (reproducidas en la **tabla 30.8**) contiene el texto "Marque todas las opciones que apliquen". Ello significa que la respuesta a dicha pregunta no

Tabla 30.7 Ejemplo de pregunta de profundización

2	¿Cómo calificaría la atención médica que ha recibido desde cuando fue trasplantado?	1	Excelente	*Pase a 4*
		2	Buena	*Pase a 4*
		3	*Regular*	*Pase a 3*
		4	*Mala*	*Pase a 3*
		5	*Muy Mala*	*Pase a 3*
		99	*No responde*	*Pase a 3*
3	¿Cuál es el principal factor que le hace sentir que no es buena la atención médica que ha recibido desde su trasplante?			
	Pase a 3			

Fuente: Elaboración propia.

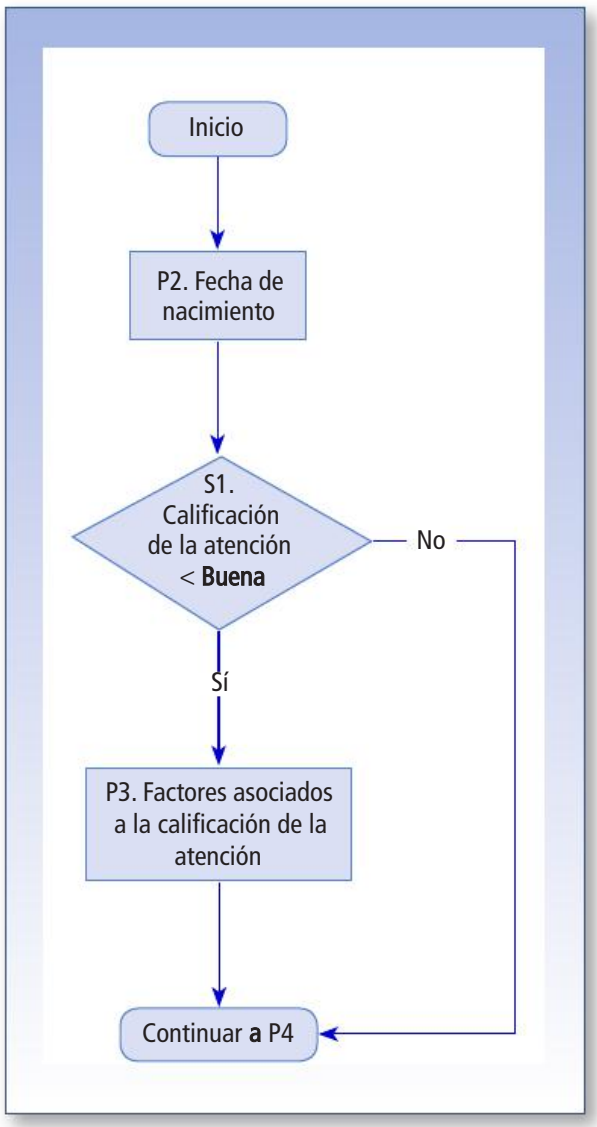


Figura 30.9 Ejemplo del diagrama de flujo correspondiente con una pregunta de profundidad.

es única, sino una combinación de respuestas, lo cual, a su vez, implica que un único valor no sería suficiente para almacenarla.

El patrón de diseño que responde a una pregunta con respuesta múltiple no se encuentra en el modelo de captura, sino en el de datos, y consiste en trasladar esta pregunta a su propia entidad, de forma que cada instancia de la entidad corresponda a una de las opciones seleccionadas. La **figura 30.10** presenta la implementación de este patrón, incluyendo el uso de una terminología médica para controlar los nombres de los medicamentos.

En el diagrama de la **figura 30.10**, la entidad que almacena las respuestas de la pregunta 4 es Medicamentos; específicamente, el atributo id_rxnorm.

EJEMPLO

Si en este punto, el lector tiene dudas sobre el nombre de este atributo, le recomiendo revisar la sección "Terminologías médicas". Así mismo, si el lector tiene dudas sobre la presencia de la entidad Relaciones_Participantes_Medicamentos, le recomiendo revisar la sección "Relaciones de muchos a muchos".

Preguntas para las que no se conocen todas las opciones de respuesta

Una exploración de las opciones de respuesta de la pregunta 4 del ejemplo **(tabla 30.8)** revela que aún oculta una sorpresa más: admite la opción de repuesta "Otro". Esta opción se explora en la pregunta de profundización 5, que, a su vez, es abierta.

No es raro que al definir un cuestionario no se tenga un listado de todas las posibles opciones de respuesta de una pregunta, como en el caso de los medicamentos prescritos a los participantes de nuestro estudio de adherencia. Muy a menudo, esto se soluciona incluyendo una opción "Otro" y una pregunta de profundización abierta. Tal opción puede ser válida en escenarios donde estos valores no hacen parte del análisis estadístico.

Lamentablemente, ese no era el caso en nuestro estudio: nosotros necesitábamos clasificar a los participantes de acuerdo con el número de medicamentos que tomaban a diario. El problema es que combinar la opción "Otro" y la pregunta abierta de profundización abría la posibilidad de que nuestros datos contuvieran más de un nombre para el mismo medicamento.

En este punto, tal vez no sea sorprendente que la solución provino de implementar el modelo de datos de la **figura 30.10**. Sin embargo, el mensaje que quisiera que se llevara el lector es que esta combinación, de una opción "Otro" más una pregunta de profundización abierta, es un negocio peligroso. En la mayoría de los casos, es mejor invertir más tiempo en definir las opciones de respuesta o apoyarse en una terminología médica.

EJEMPLO

Una nota final

El hecho de que los requerimientos impuestos por las preguntas 4 y 5 ya hubieran sido solucionados por el modelo de datos demuestra no solo mi capacidad para seleccionar ejemplos, sino también, el poder de estos métodos para definir sistemas de información flexibles que se adaptan a varias necesidades.

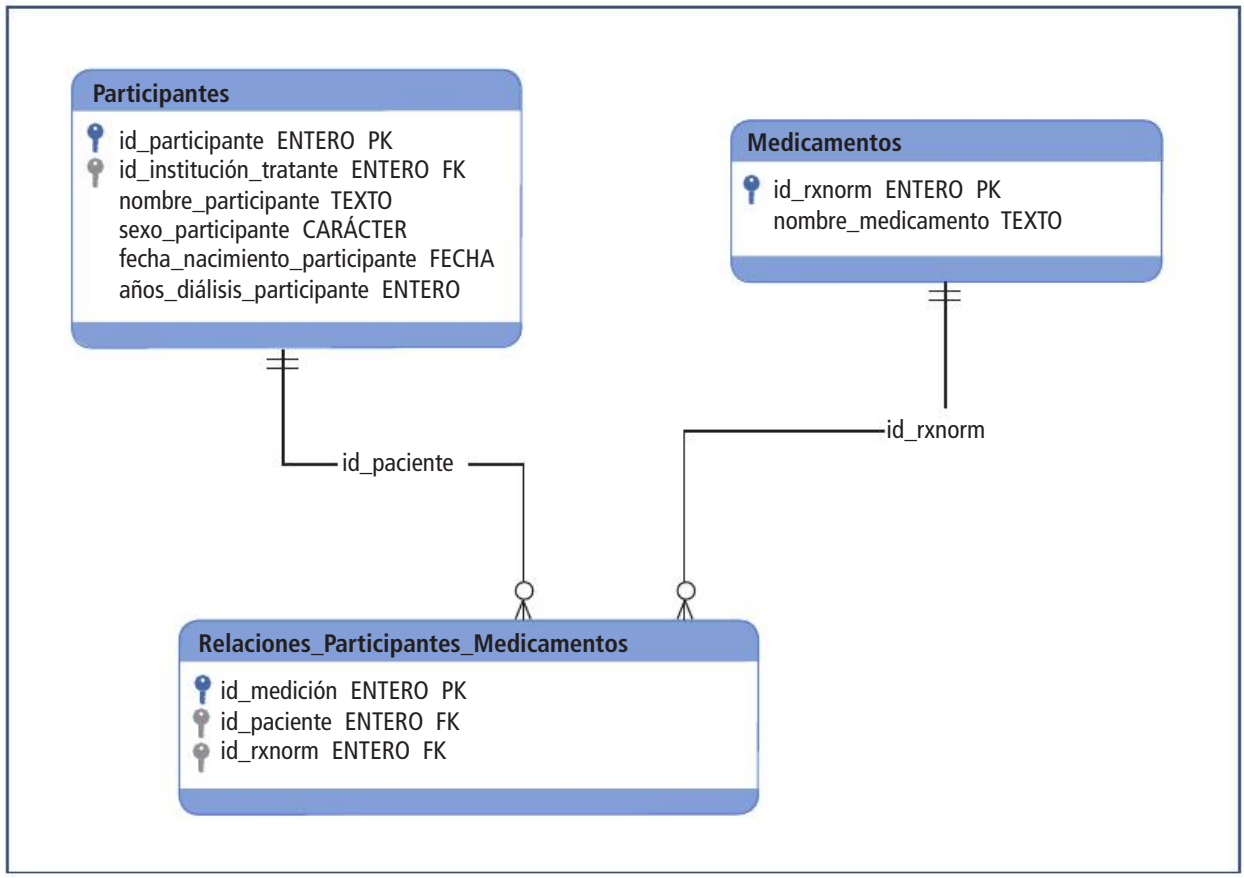


Figura 30.10 ERD correspondiente a los medicamentos que toma cada participante.

Tabla 30.8 Ejemplo de pregunta con respuesta múltiple

4	¿Qué medicamentos toma diariamente? (marque todas las opciones que apliquen)	1	Ácido micofenólico	
		2	Gemfibrozilo®	
		3	Tacrolimus	
		4	Ciclosporina	
		5	Sirolimus	
		98	*Otro*	*Pase a validación 2*

Fuente: Elaboración propia.

Referencias

1. Hoberman S. Data modeling made simple: a practical guide for business and IT professionals. 2nd ed. Bradley Beach, N.J: Technics Publications; 2009.
2. Barker R. Case Method: Entity relationship modelling, Volume 1. Michigan: Addison-Wesley; 1990.
3. Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. Med Care. 2012;50(Suppl):S60-7.
4. Harrington JL. Relational database design clearly explained. 2nd ed. New York: Morgan Kaufmann Publishers; 2002.
5. Hay DC. Data model patterns a metadata map. Amsterdam; Boston: Elsevier Morgan Kaufmann; 2006.
6. RxNorm. U.S. National Library of Medicine [internet]. 2014 [citado 2015 ene. 2]. Disponible en: http://www.nlm.nih.gov/research/umls/rxnorm/
7. World Health Organization (WHO). International

Classification of Diseases (ICD) [internet]. 2015 [citado 2015 ene. 2]. Disponible en: http://www.who.int/classifications/icd/en/
8. SNOMED-CT [Internet]. 2015 [citado 2015 ene. 2]. Disponible en: http://www.ihtsdo.org/snomed-ct
9. Medical Subject Headings, U.S. National Library of Medicine [Internet-Home Page]. 2015 [citado 2015 ene. 2]. Disponible en: http://www.nlm.nih.gov/mesh/meshhome.html
10. Logical Observation Identifiers Names and Codes (LOINC®)- LOINC [Internet]. 2015 [citado 2015 ene. 2]. Disponible en: http://loinc.org/
11. Suárez-Obando F, Camacho J. Estándares en informática médica: generalidades y aplicaciones. Rev Colomb Psiquiatr. 2013;42:295-302.



CAPÍTULO 31

Construcción de bases de datos

JHON CAMACHO

El foco del capítulo anterior estuvo en la definición de las características que deben cumplir los datos del estudio para que se les pueda considerar válidos, así como la secuencia de acciones que deben cumplirse durante su captura. Para documentar este ejercicio, se introdujeron los conceptos ***modelo de datos*** y ***modelo de captura***. En este capítulo, veremos cómo trasladar las definiciones del primero de dichos modelos al computador, en la forma de una base de datos. Nuestro objetivo en dicho proceso no se limita a construir la estructura que almacenará los datos, sino que también incluye indicarle al programa las definiciones del modelo (como en el caso de los valores permitidos), de forma que este alerte cuando se ingresen datos no válidos.

La aplicación de los conceptos del capítulo anterior es independiente de la tecnología que se use para la creación de la base de datos o para la captura. Sin embargo, ese no es el caso aquí. Hay muchas alternativas informáticas para implementar una base de datos, desde archivos de texto hasta motores de bases de datos con capacidad para almacenar billones de registros y responder miles de consultas simultáneas.

Cada opción trae sus costos y sus beneficios. En un lado del espectro, almacenar la información en archivos de texto sin estructura solo requiere un nivel básico de conocimiento, y, en general, todos los computadores de escritorio cuentan con un programa de procesamiento de texto. Sin embargo, dicha alternativa no ayuda a validar los datos durante la captura, lo que aumenta sustancialmente la probabilidad de ingresar datos no válidos. Por otra parte, una base de datos de tipo empresarial, como Oracle o MySQL, provee una miríada de herramientas para validar, transformar y consultar los datos, pero su uso requiere una curva de aprendizaje pronunciada. Eso, sin hablar del costo de la licencia, que puede ser bastante alto en casos como el de Oracle.

Una solución intermedia, que se ajusta al caso de estudios de pequeña a mediana envergadura, es usar Microsoft Excel. Los ejemplos del presente capítulo utilizan este programa.

EJEMPLO

¿Dónde trazar la línea?

La línea que separa un estudio de mediana envergadura de otros más grandes es bastante difusa. Mi recomendación es usar Excel para administrar los datos de un estudio que cumpla las siguientes características:

- La base de datos solo necesita ser accedida por un usuario a la vez.
- El número de preguntas por cuestionario no es muy grande, y el número de cuestionarios tam-

poco lo es. No hay un número mágico para la recomendación, pero, en general, piense que si el proyecto se ve grande, probablemente necesita una solución robusta.

- Los datos se capturan en papel o en cuestionarios electrónicos (se tratará más sobre esto en el siguiente capítulo), y luego se digitan manualmente en la base de datos.
- Los saltos condicionales (definidos en el modelo de captura) solo dependan de la información capturada en el mismo cuestionario.
- Los saltos condicionales no son controlados por el computador. Es decir, el recolector y el digitador son responsables de hacer cumplir las definiciones del modelo de captura.

Si el estudio en cuestión no cumple alguna de las mencionadas características, es recomendable crear un programa de captura a la medida, en cuyo caso se necesitan herramientas más potentes, que pueden ir desde un manejador de bases de datos de escritorio, como Microsoft Access o LibreOffice Base, hasta una aplicación web, basada en un manejador de bases de datos de escala empresarial, como Oracle o MySql.

TRASFERENCIA DE LAS DEFINICIONES DEL MODELO DE DATOS

Cada componente del modelo de datos sirve de plano de construcción de un elemento de la base de datos. Las entidades se convierten en tablas; los atributos, en columnas, y las relaciones… bien, las de las bases de datos también se llaman relaciones. De esta forma, lo que en la etapa de modelado (véase el **capítulo 30** "Modelado") eran conceptos abstractos ahora se convierten en componentes concretos con los que puede interactuar una persona.

El primer paso en la construcción de la base de datos es crear un archivo de Excel en blanco, lo que puede lograrse través de la pantalla de inicio **(figura 31.1)** o utilizando las opciones del menú *Archivo* ⟶ *Nuevo*.

EJEMPLO

Versiones de Excel

La imagen de la **figura 31.1** corresponde a la versión 2013 de Excel. Sin embargo, todas las instrucciones de este y del siguiente capítulo "Preparación de datos" funcionan igualmente bien en la versión 2010 y en la versión 2011 para Mac osx. Dicho esto, es posible que el lector encuentre algunas diferencias menores. En estos casos, el texto aclara en qué consiste la diferencia y cómo se debe proceder en cada versión.

Entidades

En la mayoría de las bases de datos usadas en estudios clínicos, la información se almacena en tablas.

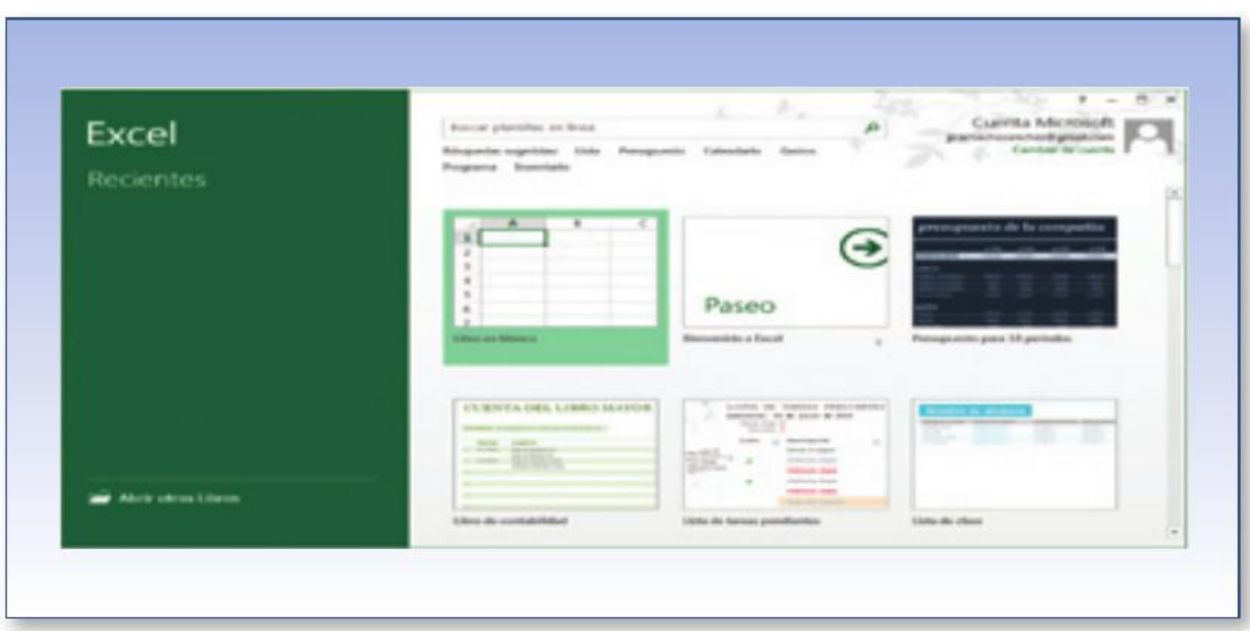


Figura 31.1 Excel-pantalla de inicio.

Las tablas constituyen colecciones de registros (o filas). Estos, a su vez, representan los objetos, los sucesos o los actores que interactúan en el proceso representado por la información.

Si dichas definiciones se parecen a las de entidad e instancia de una entidad (presentadas en el capítulo anterior), es porque las tablas y los registros constituyen la materialización de las entidades y sus instancias.

En otras palabras, las tablas de la base de datos almacenan la información de los conceptos representados por las entidades del modelo de datos. Adicionalmente, así como las entidades se componen de atributos que describen las características de cada instancia, las tablas tienen campos (o columnas) que almacenan los valores asociados a cada registro.

EJEMPLO

Libros, hojas y bases

Excel permite crear más de una tabla por hoja. Técnicamente, una hoja podría almacenar la información correspondiente a más de una entidad (de hecho, toda la base de datos podría estar contenida en una misma hoja). Sin embargo, mantener una única tabla por hoja facilita la búsqueda de información, no solo porque es más fácil navegar a la tabla con la cual se quiere interactuar, sino porque muchas de las herramientas de Excel (por ejemplo, los filtros) operan sobre toda la fila de la hoja. Es decir, si se tienen dos tablas en una hoja, una al lado de la otra, al filtrar una de ellas, se ocultan filas de la otra. Ubicar una tabla debajo de la otra tampoco es práctico, pues para acceder a los datos de una tabla habría que recorrer todos los registros de las tablas ubicadas encima. En definitiva, la mejor opción es *una tabla* por hoja.

En cuanto a la relación entre base y libro, Excel solo permite crear relaciones entre tablas del mismo libro. Ello, teóricamente, permitiría separar la base de datos en más de un archivo, siempre y cuando estos contuvieran tablas no relacionadas. Sin embargo, hacerlo puede conducir a problemas de manejo de versiones. Tener toda la base en un solo archivo significa que cualquier cambio, en cualquiera de las tablas, altera la fecha de modificación del archivo, lo cual facilita verificar la versión de la base de datos con la que se está trabajando en un momento dado. Si, por el contrario, la base se aloja en más de un archivo, para conocer la fecha de la última modificación sería necesario comparar las fechas de todos los archivos. De nuevo, la mejor opción es "1 a 1", un archivo por base.

La **figura 31.2** presenta la entidad *Participantes* como se la definió en el **capítulo 30** "Modelado". Para construir la tabla correspondiente, se pueden desarrollar los siguientes pasos:

1. **Inserte una nueva (*Shift* + *F11*). Si el libro ya** tiene una hoja en blanco (como cuando acaba de ser creado), puede saltar este paso.
2. **Cambie el nombre de la hoja para que co**rresponda al nombre de la entidad (***Participantes***). **Para cambiar el nombre de la hoja,** puede hacer clic con el botón secundario del ratón y seleccionar ***Cambiar nombre***.
3. **Seleccione la celda A1 y utilice la opción** *Insertar* → *Tablas* → *Tabla*. Dependiendo de la versión de Excel que esté usando, puede que aparezca una ventana como la que se muestra en la **figura 31.3**, en cuyo caso se debe hacer clic en el botón *Aceptar*.
   El resultado debe ser una tabla con un único campo llamado ***Columna1***, como la que se muestra en la **figura 31.4**.
4. **En la primera fila de la tabla, ingrese los** encabezados correspondientes a los atributos de la entidad. El resultado debería verse como la tabla de la **figura 31.5**.
   En la **figura 31.5, puede llamar la atención** que el encabezado correspondiente al atributo ***años_diálisis_participante*** **se introdujo**



Figura 31.2 Entidad Participantes.

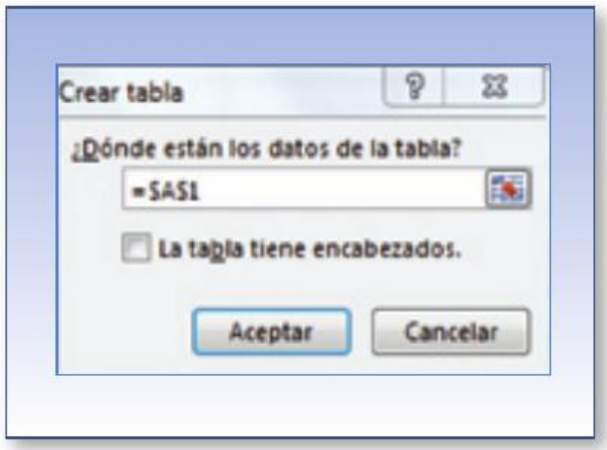


Figura 31.3 Confirmación de creación de tabla.

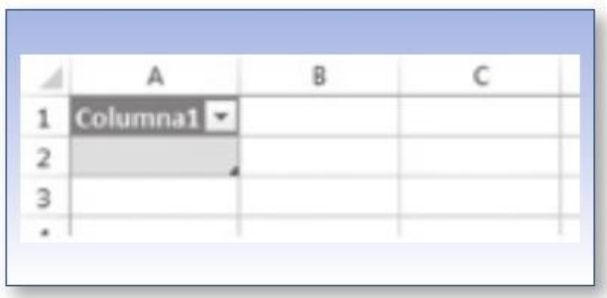


Figura 31.4 Nueva tabla.

como *anos_dialisis_participante*. En general, es recomendable no incluir "ñ", ni tildes ni diéresis en los nombres de las tablas o de los campos. La razón para ello es que dichos nombres van a aparecer en los datos cuando se extraigan para enviarlos al programa de análisis, y en ausencia de la configuración necesaria (que suele no ser automática), estos programas pueden tener problemas para interpretar tales símbolos.

5. **Haga clic en cualquiera de las celdas de la** tabla (como se ve en la **figura 31.5, en este** punto, la tabla debe cubrir las celdas de la A1 a la E2).

6. **En Excel 2010 y 2013: vaya a la cinta de** comandos ***Herramientas de tabla*** ⟶ *Diseño* **y al lado izquierdo en el campo** *Nombre de la tabla* **ingrese *Participantes*.** En Excel para Mac 2011: vaya a la cinta de comandos Tablas, y haga clic en Cambiar nombre. Esto debe activar una caja de texto a la izquierda de la barra de fórmulas (resaltada en la **figura 31.6**). **Reemplace el nombre** temporal ("Tabla 4" en la **figura 31.6**) **por** *Participantes*.

En este punto, Excel reconoce las celdas de la A1 a la E2 (resaltadas en la **figura 31.7**) como pertenecientes a la tabla ***Participantes***. **El área** cubierta por la tabla está demarcada por la celda que sirve de encabezado del primer campo y la última celda de datos del último campo (marcada con una punta azul oscura en la parte inferior derecha de la celda E2 de la **figura 31.7**).

Versiones de Excel: Campos y columnas
Excel solo reconoce como pertenecientes a la tabla los datos que se encuentran dentro del área de ella. Esto significa que un valor ubicado en la celda D4 (de acuerdo con el ejemplo de la **figura 31.7**) no pertenece a la tabla. Para Excel, este dato pertenece a la columna D, pero no a la tabla *Participantes*. Para llamar la atención sobre eso, a lo largo de este y del siguiente capítulo, cuando me refiero a las celdas ubicadas dentro de la tabla, uso el término *campo*, mientras que cuando me refiero a las celdas, independientemente de si pertenecen o no a la tabla, uso el término *columna*.

En adelante, Excel ampliará automáticamente esta área cuando se incluyan nuevos campos o registros, siempre y cuando estos se incluyan dentro o en las celdas adyacentes al área actual. O sea,



Figura 31.5 Tabla *Participantes*.

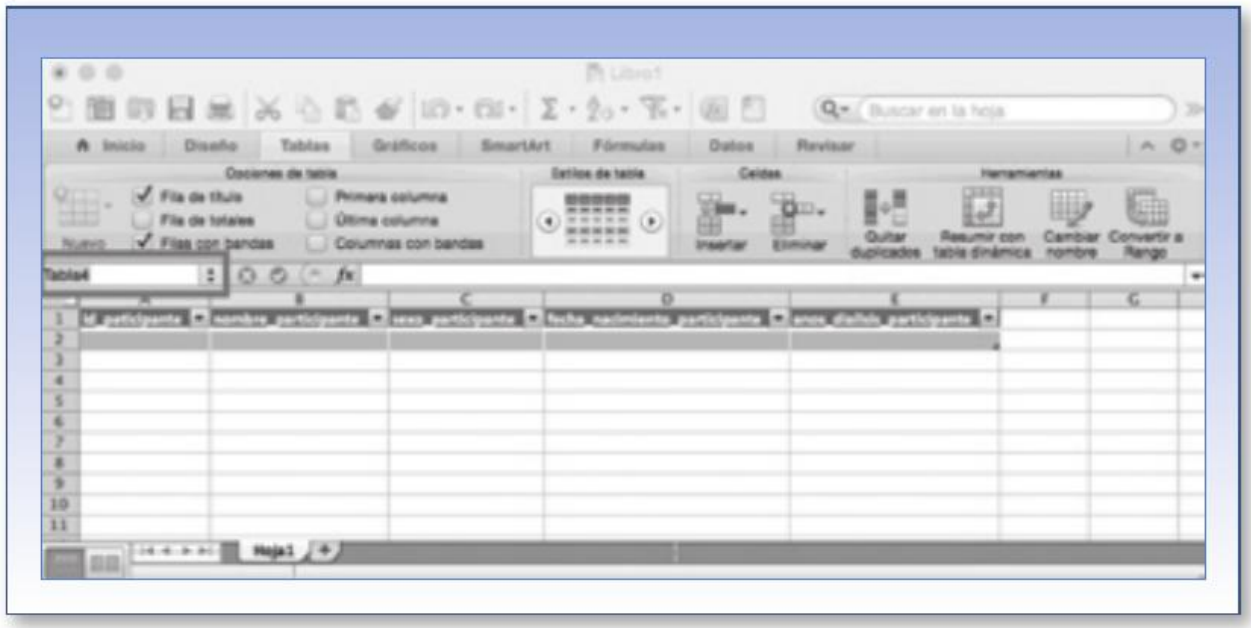


Figura 31.6 Cambio del nombre de una tabla en Excel para Mac 2011.



Figura 31.7 Área de la tabla.

para ampliar la tabla basta con ingresar nuevos encabezados o nuevos registros, siempre y cuando no se dejen columnas o filas en blanco. Para ver esto, cree el campo *id_institucion_tratante* (ingresando el encabezado en la celda F1), e ingrese los registros que se muestran en la **figura 31.8**. Para dar formato o borrar el contenido de las celdas de la tabla, puede usar las mismas herramientas que usaría con cualquier otra celda de la hoja.

Al usuario interesado en profundizar en el manejo de tablas en Excel, además de otros conceptos presentados en este capítulo, le recomiendo consultar ***Excel 2013 Bible***, de John Walkenbach[1].

Valores permitidos

El siguiente paso en la construcción de la base de datos consiste en indicarle a Excel cómo validar que los datos ingresados en la tabla cumplan las definiciones de valores permitidos especificadas en el modelo de datos.



La herramienta de validación de datos de Excel permite definir reglas que especifican qué valores aceptar en cada campo.



Figura 31.8 Ingreso de nuevos campos y registros.

Una vez definidas dichas reglas, Excel las utilizará para prevenir que el usuario ingrese valores no permitidos, como se ve en la **figura 31.9**.

Para especificar los valores permitidos de un campo, se pueden llevar a cabo los siguientes pasos:

1. Seleccione las celdas del campo sin incluir el encabezado.
2. Seleccione *Datos* ⟶ *Herramientas de datos* ⟶ *Validación de datos*. Debe aparecer una ventana como la que se muestra en la **figura 31.10**.
3. Seleccione la pestaña *Configuración*.
4. Seleccione una opción de la lista *Permitir*. Cada opción permite especificar un tipo de criterio de validación y requiere una configuración.
5. Si el campo corresponde a un atributo obligatorio, desmarque la casilla *Omitir blancos*.
6. (opcional) Seleccione la pestaña *Mensaje de entrada*. En esta pestaña se puede configurar un mensaje informativo, el cual Excel presentará cada vez que se seleccione una celda del campo. Este mensaje puede usarse para explicarle al usuario cuáles valores se permiten para este campo.
7. Seleccione la pestaña *Mensaje de error*.
8. En la lista *Estilo*, seleccione *Detener*. Eso informará a Excel que no debe permitir que el usuario ingrese un valor que transgreda el criterio de verificación especificado en el punto 4.
9. En la caja de texto *Mensaje de error*, ingrese un mensaje que indique los valores permitidos para el campo. Excel presentará dicho mensaje cada vez que se intente ingresar un valor no permitido.
10. Haga clic en el botón *Aceptar*.

Las siguientes secciones explican cómo utilizar las opciones del paso 4 para especificar diferentes tipos de validaciones.

Validación de datos numéricos, fechas y horas

Para especificar el rango permitido de un dato numérico, se puede seleccionar la opción *Número entero*, o la opción *Decimal*, dependiendo de si el campo corresponde a un atributo de tipo entero o de punto flotante (números reales con parte decimal). En la lista *Datos*, seleccione la opción que corresponda al tipo de verificación que quiera aplicar al valor numérico **(figura 31.11)**.

De acuerdo con su selección, deben aparecer cajas de texto en las puede ingresar los valores que se usarán para aplicar la regla de validación. Por ejemplo, la **figura 31.12** muestra la configuración de una regla en la cual se exige que el valor del campo sea > 1 y < 2,5.

El procedimiento es prácticamente igual cuando se trata de fechas o de horas. La **figura 31.13** muestra un ejemplo para un campo de tipo fecha, mientras que la **figura 31.14** muestra un ejemplo para un campo hora.

Al especificar estos criterios, es importante tener en cuenta que las validaciones *Entre* y *No está entre* incluyen en el intervalo los valores límite. Es decir, si se especifica que el valor debe estar entre 1 y 5, entonces 1 y 5 son valores permitidos. Similarmente, si se especifica que el valor no puede estar entre 1 y 5, entonces ni 1 ni 5 serán aceptados. Adicionalmente, el formato en el que se deben introducir las fechas depende de la configuración de cada computador; en particular, en lo que se refiere al orden del mes y el día (31/01/2015 *vs.* 01/31/2015). Una forma

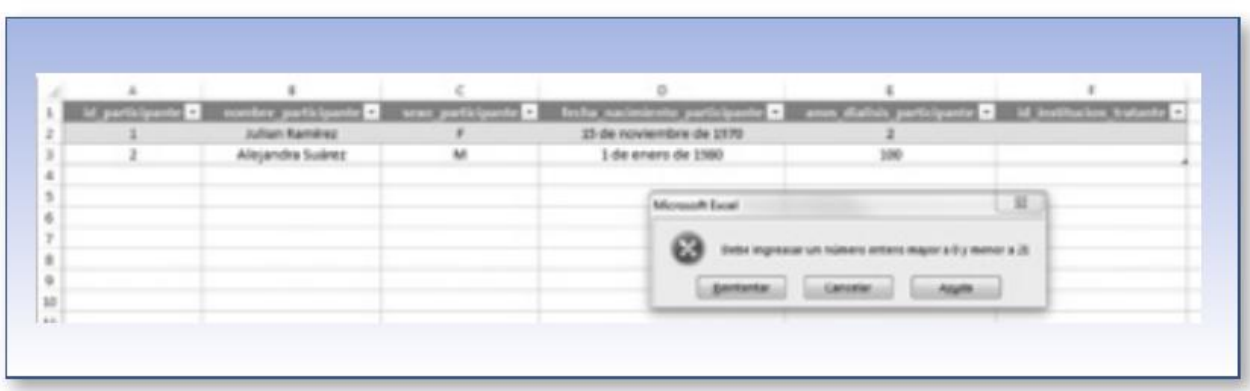


Figura 31.9 Cómo prevenir el ingreso de valores no permitidos.

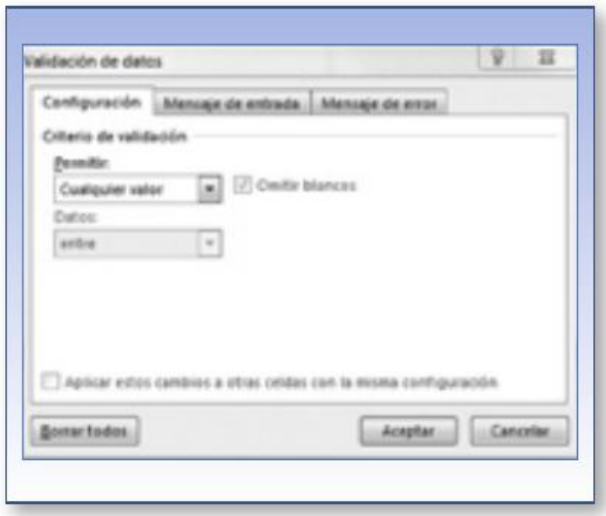


Figura 31.10 Configurando la validación de datos.

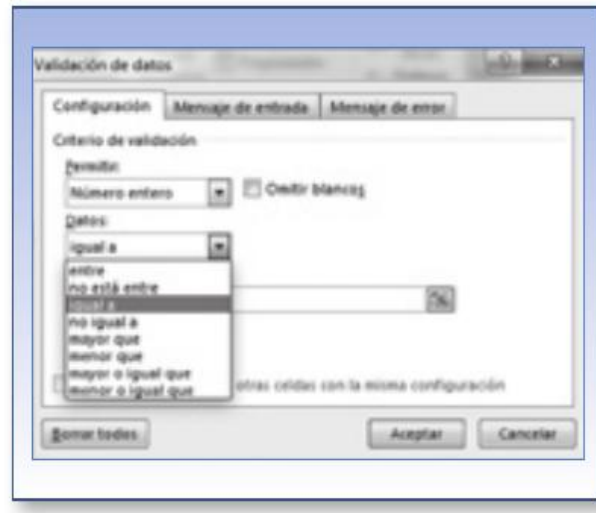


Figura 31.11 Cómo especificar la validación de datos numéricos.

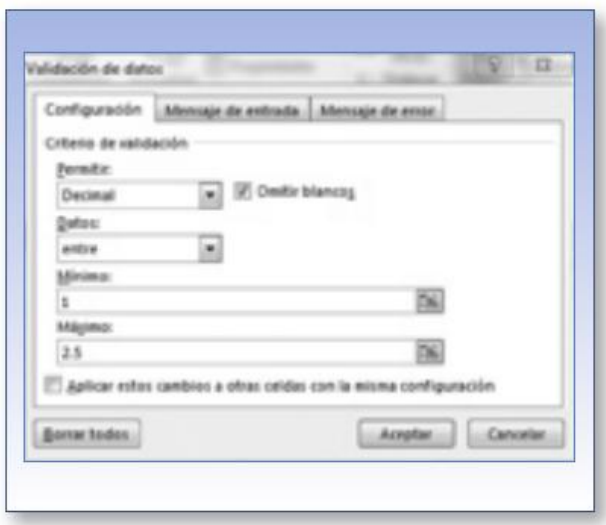


Figura 31.12 Especificación de una validación *Entre* para un campo de tipo numérico.

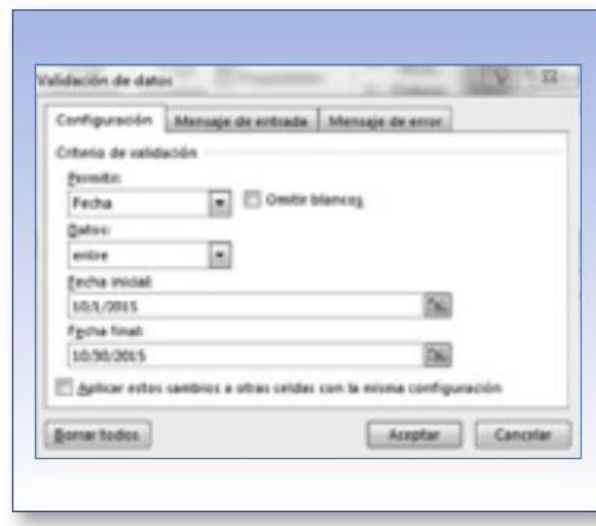


Figura 31.13 Especificación de una validación para un campo de tipo fecha.

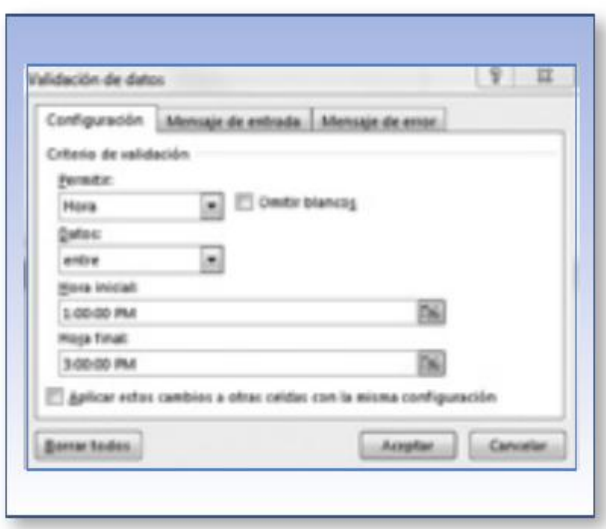


Figura 31.14 Especificación de una validación para un campo de tipo hora.

práctica para saber qué formato debe usarse es ingresar una fecha con un día del mes igual a 13 o mayor, como 20/12/2001. Si Excel no acepta esa fecha, se puede probar invirtiendo el mes y el día (12/20/2001).

Validación de la longitud de un texto

Si la definición de valores permitidos de un atributo de tipo texto solo se refiere a la longitud del valor (sin tener en cuenta su significado), se puede especificar la regla de validación usando la opción ***Longitud del texto***, cuyos parámetros son muy similares a los empleados para la validación de datos numéricos **(figura 31.15)**.

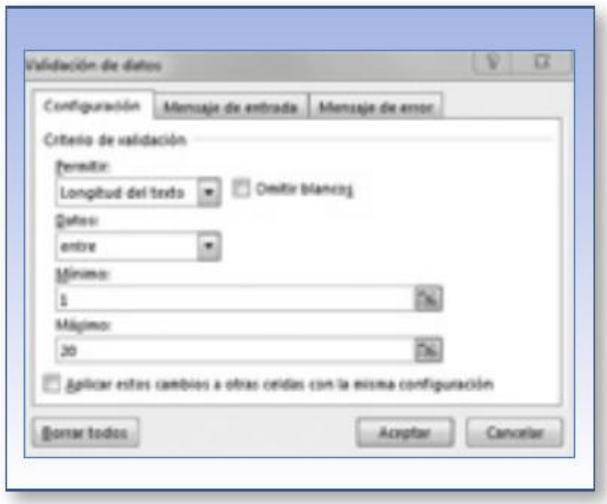


Figura 31.15 Especificación de la validación de la longitud de un texto.

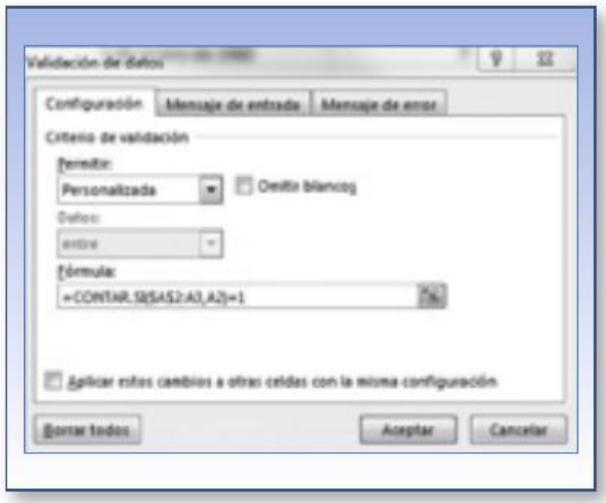


Figura 31.16 Especificación de la validación de un campo con valores únicos.

Validación de identificadores y otros valores únicos

Algunos campos, como los que corresponden a atributos identificadores, deben cumplir la regla de que sus valores nunca se repitan. Para especificar este criterio de validación, se puede usar la opción ***Personalizada***, como se muestra en la **figura 31.16**.

El ejemplo anterior asume que el campo se ubica en la columna "A"; en caso contrario, sería necesario reemplazar todas las "A" de la fórmula para que correspondieran con la columna del campo. Adicionalmente, la formula asume que el último registro de la tabla está ubicado en la fila 3; en caso contrario, sería necesario reemplazar este número para que coincida con la fila del último registro; o por el número 2, en caso de que la tabla se encuentre vacía. Estos reemplazos solo son necesarios cuando se define la regla por primera vez, pues de ahí en adelante Excel se encarga de actualizar la fórmula para que funcione con los nuevos registros.

Validación de datos a partir de una lista fija de valores

En algunos casos, la definición de valores permitidos de un atributo consiste en una lista fija de valores, como en el caso del atributo *sexo_participante* **(tabla 31.1)**.

La validación de estos datos se puede especificar usando la opción ***Lista***, como se muestra en la **figura 31.17**.

La casilla de verificación ***Celda con lista desplegable*** le dice a Excel que debe presentar la lista de los valores permitidos cuando se entra a una de las celdas del campo **(figura 31.18)**. El usuario puede digitar el valor directamente o seleccionar uno de la lista.

Validación de valores de referencia y otras relaciones

La validación basada en listas de valores se puede llevar un paso más allá cuando el modelo de datos incluye entidades que agrupan valores de referencia (véase "Valores de referencia", en el **capítulo 30** "Modelado"). La **figura 31.19** muestra un ejemplo de esta técnica, en el que se define que la entidad ***Instituciones_Tratantes*** agrupa los valores permitidos para el atributo *id_institución_tratante*, que, a su vez, almacena la información del centro de atención relacionado por un participante.

Para transferir esta definición a la base de datos, se pueden desarrollar los siguientes pasos:

Tabla 31.1 Definición de valores para el atributo sexo_participante

Atributo	Tipo de dato	Valores	Formato	Significado	Obligatorio
sexo_participante	CARÁCTER	M, F		*M = Masculino* *F = Femenino*	Sí

Fuente: Elaboración propia.

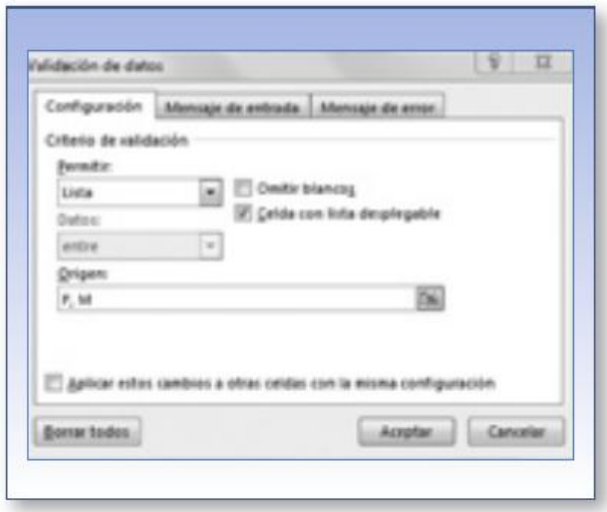


Figura 31.17 Especificación de la validación de una lista de valores.

1. Cree la tabla *Instituciones_Tratantes* e ingrese algunos datos, como lo muestra la **figura 31.20**. No olvide cambiar el nombre de la hoja por *Instituciones_Tratantes*.
2. Cree un nuevo campo llamado *seleccion*.
3. Seleccione la primera celda de datos del campo *seleccion* (C2) e ingrese la fórmula:

```
=CONCATENAR([@[id_institucion_tratante]],” - “,
    [@[nombre_institucion_tratante]])
```

   Para facilitar el ingreso de la fórmula, las referencias a los campos (como *[@[id_institucion_tratante]]*) se pueden obtener haciendo clic en la primera celda de datos de cada campo. El resultado debe parecerse al que muestra la **figura 31.21**.
4. Seleccione las celdas de datos del campo *id_institución_tratante*, de la tabla *Participantes*, e ingrese los criterios de validación que se muestran en la **figura 31.22**.
5. En la caja de texto *Origen*, ingrese la siguiente formula reemplazando “C” por la letra correspondiente a la columna en la que se creó el campo *selección* durante el paso 2:

```
=Instituciones_Tratantes!$C$2:$C$100
```

6. Haga clic en el botón *Aceptar*.

En adelante, cada vez que se ingresen o se eliminen registros en la tabla *Instituciones_Tratantes*, Excel actualizará automáticamente la lista de valores permitidos para el campo *id_institución_tratante*. La fórmula del paso 5 asume que la tabla con los valores de referencia no tiene más de 100 registros. En caso contrario, se debe ampliar este número.

El resultado en la tabla *Participantes* debe parecerse al que se muestra en la **figura 31.23**.

A pesar de que esta técnica es comúnmente usada con valores de referencia, en realidad puede emplearse con cualquier relación *uno-a-muchos* entre entidades del modelo de datos. Lo importante es recordar que el campo de selección debe crearse concatenando el atributo identificador y un atributo de descripción de la entidad que está en el lado *muchos*, y que la regla de validación debe especificarse en el atributo que sirve de referencia foránea en la entidad que se ubica en el lado *uno*.

Validación de otros datos obligatorios

Cuando el campo corresponde a un atributo obligatorio y, además, requiere una validación como las que se presentan en las secciones anteriores, basta con desmarcar la casilla *Omitir blancos*, para que Excel aplique la regla a las casillas vacías, lo que en la mayoría de los casos significa que el usuario tendrá que ingresar un valor antes de continuar. Sin embargo, ¿qué pasa cuando el único criterio de validación es que el campo no debe quedar vacío? En esos casos, se puede usar la opción *Personalizada*, como se muestra en la **figura 31.24**, y se selecciona *Omitir blancos*.



Figura 31.18 Lista de valores permitidos presentada durante el ingreso de los datos.

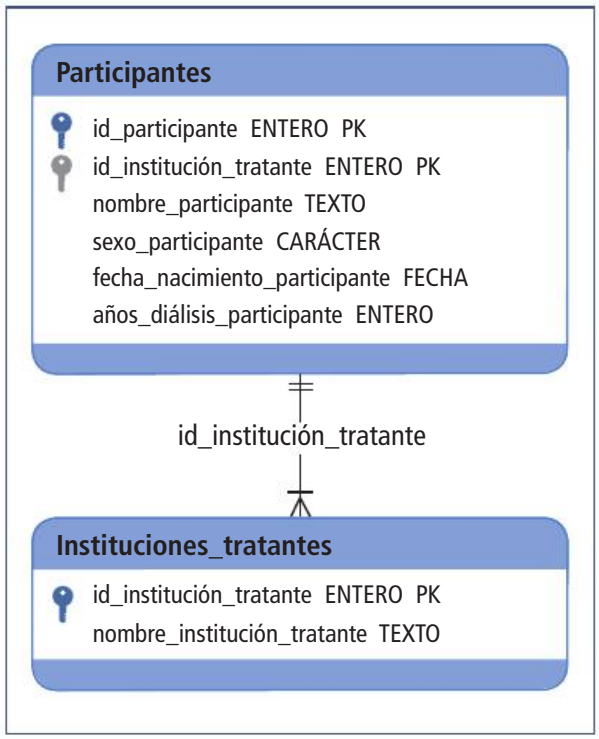


Figura 31.19 Ejemplo de la inclusión de valores de referencia.

El ejemplo anterior asume que el campo se ubica en la columna "B"; en caso contrario, sería necesario reemplazar la "B" de la fórmula para que correspondiera con la columna del campo. No olvide desmarcar la casilla ***Omitir blancos***.

Validaciones personalizadas

Como lo muestran varios de los ejemplos anteriores, la opción ***Personalizada*** puede usarse para crear reglas de validación especiales. El truco consiste en ingresar una fórmula cuyo resultado sea ***FALSO*** para los valores que se quiere rechazar. Como lo muestra la **figura 31.24**, la fórmula se define como si operara solo sobre la primera celda del campo (B2, en el caso del ejemplo). Sin embargo, Excel se encarga de alterarla para que funcione con todas las celdas que se seleccionaron durante el primer paso (ver **sección** Valores permitidos) y para los nuevos registros.



	A	B
1	id_institucion_tratante	nombre_institucion_tratante
2	1	Hospital Universitario San Ignacio
3	2	Hospital de San José
4		

Figura 31.20 Tabla *Instituciones_Tratantes*.

Validación de datos ya ingresados

Las validaciones de las secciones anteriores tienen por objetivo prevenir que se ingresen datos que violen las definiciones del modelo de datos. Pero, ¿qué pasa si el campo ya tiene información en el momento en el que se especifica la regla de validación? Es posible que en tales casos la tabla contenga datos que cumplen las reglas de validación que se acaban de definir.

Para identificar rápidamente los datos que no cumplen los criterios de validación, se puede utilizar la función ***Datos*** → ***Validación de datos*** → *Rodear con un círculo datos no válidos*. La **figura 31.25** presenta un ejemplo del resultado de esta función.

EJEMPLO

Una última nota acerca de copiar y pegar

Las herramientas de validación de Excel tienen un defecto que permite ingresar información no válida, aun cuando se hayan especificado reglas de validación. El problema ocurre cuando en vez de teclear un dato, este se pega desde el portapapeles. La situación es todavía más grave cuando se copia información desde una celda que no usa reglas de validación y se pega en una que sí las usa. En esos casos, Excel no solo permite que se ingresen datos no válidos, sino que también borra la regla de validación de la celda donde se pega el dato. Debido a ello, es muy importante no usar la función de copiar y pegar cuando se ingresa información a una base de datos en Excel.



	A	B	C
1	id_institucion_tratante	nombre_institucion_tratante	seleccion
2	1	Hospital Universitario San Ignacio	1 - Hospital Universitario San Ignacio
3	2	Hospital de San José	2 - Hospital de San José

Figura 31.21 Creación de un campo de selección.

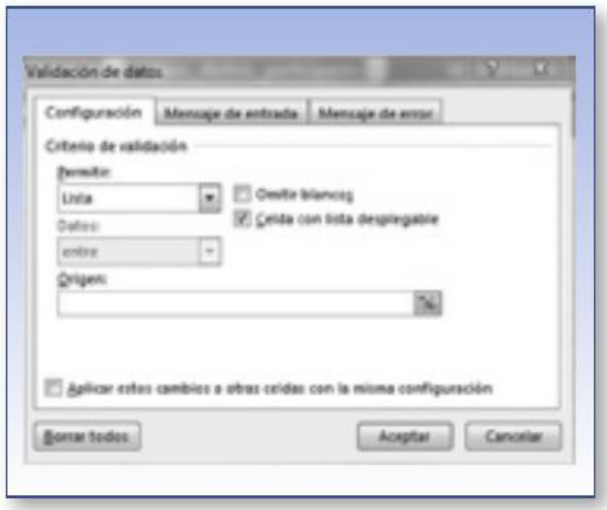


Figura 31.22 Especificación de criterio de validación basados en valores de referencia.

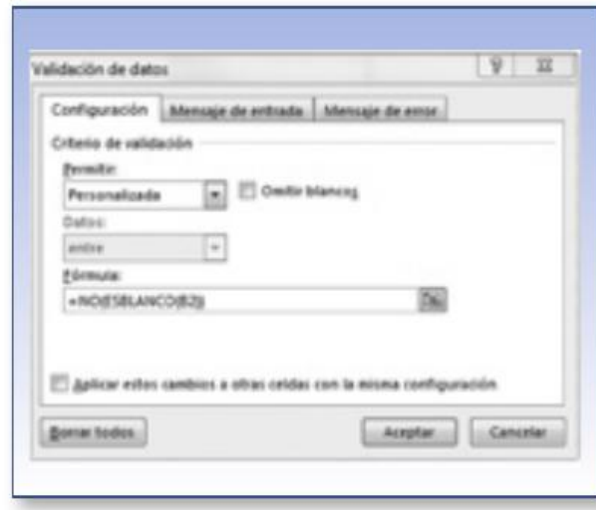


Figura 31.24 Especificación de la validación de otros campos obligatorios.



Figura 31.23 Lista de valores permitidos originada en una tabla de valores de referencia.
Fuente: Elaboración propia.



Figura 31.25 Identificación de datos que no cumplen las reglas de validación.

Referencia

1. Walkenbach J. Excel® 2013 bible. Indianapolis, IN: John Wiley & Sons; 2013.

CAPÍTULO

32 Preparación de datos

JHON CAMACHO

INTRODUCCIÓN

Las técnicas presentadas en los **capítulos 30** "Modelado" y **31** "Construcción de bases de datos" tienen por objetivo minimizar la probabilidad de que se introduzca información no válida a la base de datos del estudio. Como resultado, la representación y la organización de los datos en la base facilitan la lectura y el ingreso de los valores, pero no necesariamente facilitan su análisis con un paquete estadístico. Adicionalmente, no es recomendable importar los datos directamente al paquete estadístico desde el archivo de Excel, pues aun cuando varios de dichos paquetes cuentan con opciones para importar los datos, no es raro que los mencionados programas tengan problemas para interpretar datos provenientes de fórmulas o cuya estructura incluya más de una tabla.

En este capítulo se introducirán algunas técnicas de preparación de datos, cuyo objetivo es transformar la información de la base en una estructura más amigable para un paquete estadístico. Al final del capítulo se explica cómo exportar los datos a un formato estándar que pueda ser leído por el programa de análisis.

¿POR QUÉ ES NECESARIO TRANSFORMAR LOS DATOS?

En general, los paquetes estadísticos son diseñados para trabajar con variables que almacenan cantidades o códigos que identifican categorías. En cambio, la base de datos de un estudio debe almacenar toda la información relevante, lo que incluye no solo los valores numéricos, sino también, las descripciones de los elementos y los datos cualitativos. Así mismo, las bases de datos se diseñan de forma que la ubicación de cada variable sea única, buscando que no aparezcan ambigüedades en el momento de actualizar un valor (se profundizó sobre esto en el **capítulo 30** "Modelado"). Por el contrario, las tablas usadas en los paquetes estadísticos se organizan de foma que, en la medida de lo posible, contengan todos los datos necesarios para el análisis que se intenta realizar, lo cual busca minimizar tanto la complejidad de los comandos necesarios para el análisis como el tiempo que toma obtener los resultados.

Preparar los datos consiste en transformarlos modificando su estructura y su representación, para que se ajusten a las necesidades del análisis estadístico, y así dejar atrás la estructura y la re-

presentación con las cuales se construyó la base de datos. Dichas transformaciones son necesarias para ajustar los datos a los requerimientos de los paquetes estadísticos y para reducir los requerimientos de cómputo durante el análisis.

Tomemos, por ejemplo, el diagrama de entidad-relación de la **figura 32.1**. El modelo incluye las entidades *Participantes*, *Intituciones_Tratantes* y *Ciudades* de nuestro estudio ejemplo, acerca de la adherencia al tratamiento de pacientes con trasplante renal.

Aplicando las técnicas del **capítulo 31** "Construcción de bases de datos", el modelo de la **figura 32.1** puede implementarse en las tablas de la **figura 32.2**, la **figura 32.3** y la **figura 32.4**.

Estas tablas contienen toda la información acerca de las tres instituciones y cumplen la regla de que la ubicación de cada variable es única. Ahora, imagine que uno de los análisis estadísticos consiste en calcular la media de los años que los pacientes pasaron en diálisis (*anos_dialisis_participante*), discriminando de acuerdo con la ciudad donde se ubica su institución tratante (*id_ciudad*). Si lleváramos los datos al paquete estadístico en su organización actual, pasarían dos cosas: 1) sería necesario combinar las tres tablas en una sola antes de correr el análisis, y 2) se utilizaría espacio de almacenamiento en variables que no son útiles para el paquete estadístico, como nombre_participante. Algunos paquetes estadísticos pueden resolver el punto 1, aunque, en general, dichos programas no están diseñados para ello; en cuanto al punto 2, en circunstancias normales, cuantas más columnas tengan las tablas, más tiempo tomará extraerlas de la base de datos e importarlas al paquete estadístico. Adicionalmente, en estudios con tamaños de muestra grandes o con un numero abultado de mediciones, ahorrar espacio de almacenamiento suele ser una prioridad.

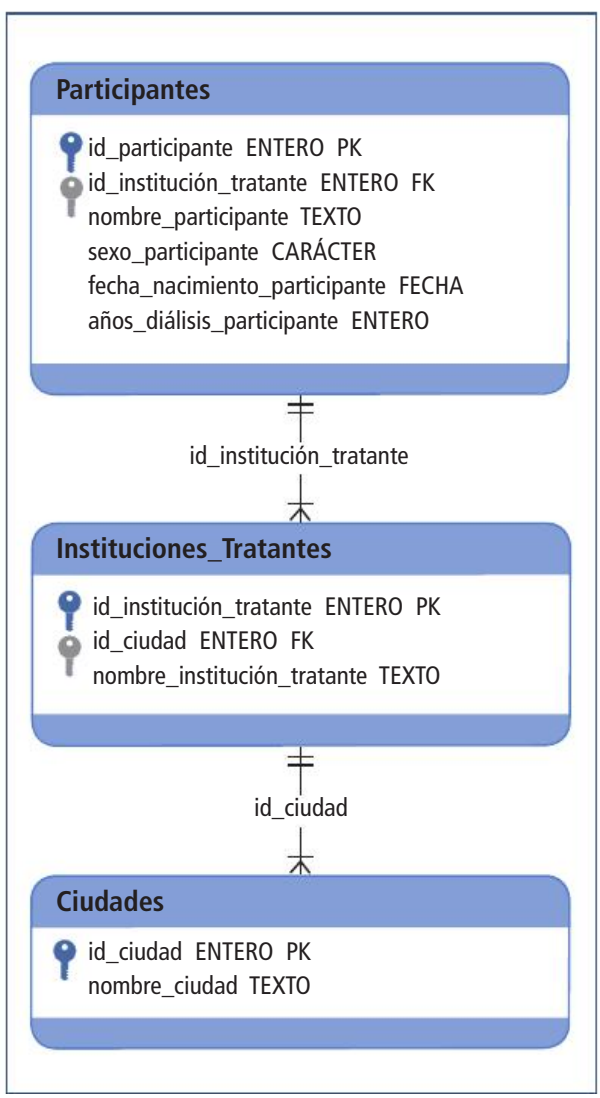


Figura 32.1 Diagrama de entidad-relación que representa a los participantes, las instituciones tratantes y las ciudades del estudio de ejemplo.

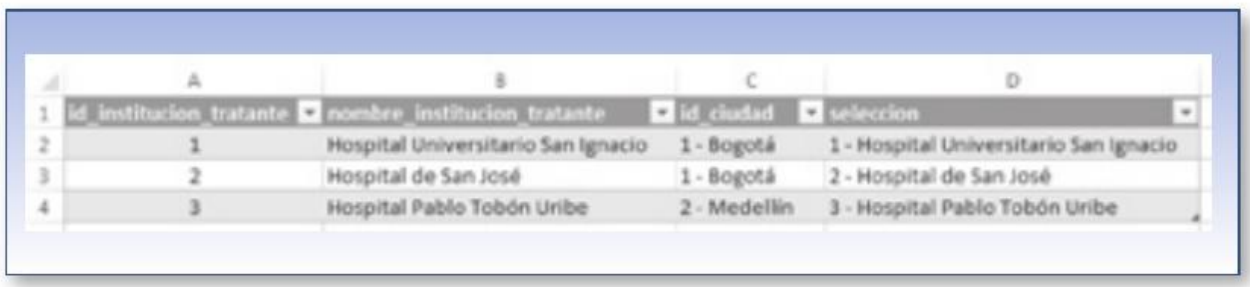
Una alternativa mejor sería construir y exportar una tabla que combinara los datos necesarios y reemplazara los valores alfanuméricos por códigos numéricos, como la que se muestra en la **figura 32.5**. Las siguientes secciones presentan las técnicas utilizadas para construir esa tabla.

CONSTRUCCIÓN DE LA ESTRUCTURA DE LA TABLA PARA EXPORTAR

El primer paso a la hora de construir la tabla que contendrá los datos para exportar es identificar la tabla origen de los registros. Cada registro deberá corresponder con una unidad del análisis que se busca desarrollar. En el caso del ejemplo, la unidad de análisis es un participante; por tanto, la tabla de origen es *Participantes*. Eso significa que cada registro de la tabla que se va a exportar debe corresponderse con un registro de la tabla *Participantes*. Así mismo, los campos de la tabla para exportar deberán corresponder con campos de la tabla *Participantes* o con campos de otras tablas que estén relacionadas directa o indirectamente con la tabla *Participantes* (en el caso del ejemplo, las tablas *Instituciones_Tratantes* y *Ciudades*).



Figura 32.2 Tabla *Participantes*, del estudio de ejemplo.



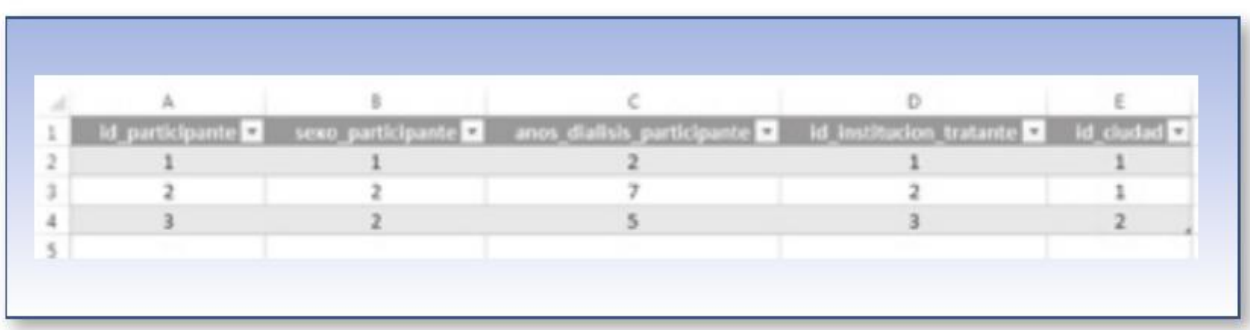
	A	B	C	D
1	id_institucion_tratante	nombre_institucion_tratante	id_ciudad	seleccion
2	1	Hospital Universitario San Ignacio	1 - Bogotá	1 - Hospital Universitario San Ignacio
3	2	Hospital de San José	1 - Bogotá	2 - Hospital de San José
4	3	Hospital Pablo Tobón Uribe	2 - Medellín	3 - Hospital Pablo Tobón Uribe

Figura 32.3 Tabla *Instituciones_Tratantes*, del estudio de ejemplo.



	A	B	C
1	id_ciudad	nombre_ciudad	seleccion
2	1	Bogotá	1 - Bogotá
3	2	Medellín	2 - Medellín
4	3	Cali	3 - Cali

Figura 32.4 Tabla *Ciudades del estudio* de ejemplo.



	A	B	C	D	E
1	id_participante	sexo_participante	anos_dialisis_participante	id_institucion_tratante	id_ciudad
2	1	1	2	1	1
3	2	2	7	2	1
4	3	2	5	3	2
5					

Figura 32.5 Tabla *Particip_Inst_Ciud*.

A continuación, se mostrará cómo para crear la estructura de la tabla que se va a exportar se siguen los mismos pasos que se emplearían para crear cualquier otra tabla de la base de datos (igualmente, se trató esto en el **capítulo 31** "Construcción de bases de datos"):

1. **Inserte una nueva hoja (*Shift + F11*).**
2. **Cambie el nombre de la hoja para que corresponda con el nombre de la tabla.** Para el ejemplo, utilizaremos ***Particip_Inst_Ciud.*** Habría sido más apropiado ***usar Participantes_Instituciones_Ciudades*, pero Excel** impone un tamaño máximo de 30 caracteres para el nombre de una hoja.

3. Seleccione la celda A1 y utilice la opción *Insertar* → Tablas → Tabla. Dependiendo de la versión de Excel que esté usando, puede aparecer una ventana como la que se muestra en la **figura 32.6**, en cuyo caso se debe hacer clic en el botón Aceptar.
   El resultado debe ser una tabla con un único campo, llamado ***Columna1***, como la que se muestra en la **figura 32.7**.

   En la primera fila de la tabla, ingrese los encabezados correspondientes a los campos. El resultado debería verse como la tabla de la **figura 32.8**.

4. Haga clic en cualquiera de las celdas de la tabla (como se ve en la **figura 32.8**, en este punto, la tabla debe cubrir las celdas de la A1 a la F2).
5. En Excel 2010 y 2013, vaya a la cinta de comandos Herramientas de tabla → Diseño y al lado izquierdo, en el campo Nombre de la tabla ingrese Particip_Inst_Ciud. En Excel para Mac 2011, vaya a la cinta de comandos Tablas, y haga clic en Cambiar nombre. Esto debe activar una caja de texto a la izquierda de la barra de fórmulas (resaltado en la **figura 32.9**). Reemplace el nombre temporal ("Tabla 3" en la **figura 32.9**) por *Particip_Inst_Ciud*.

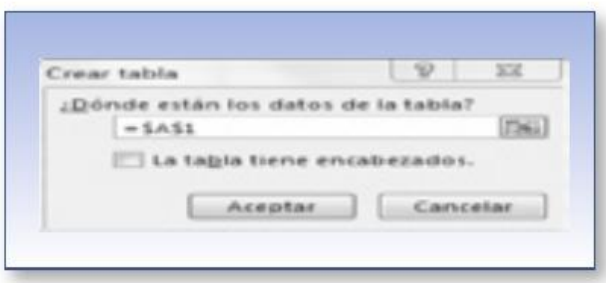


Figura 32.6 Confirmación de creación de tabla.

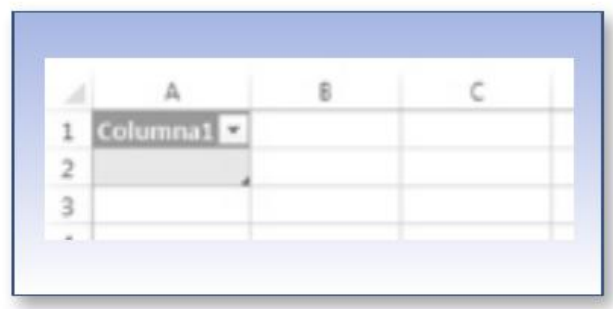


Figura 32.7 Nueva tabla.



Figura 32.8 Encabezados de la tabla *Particip_Inst_Ciud*.

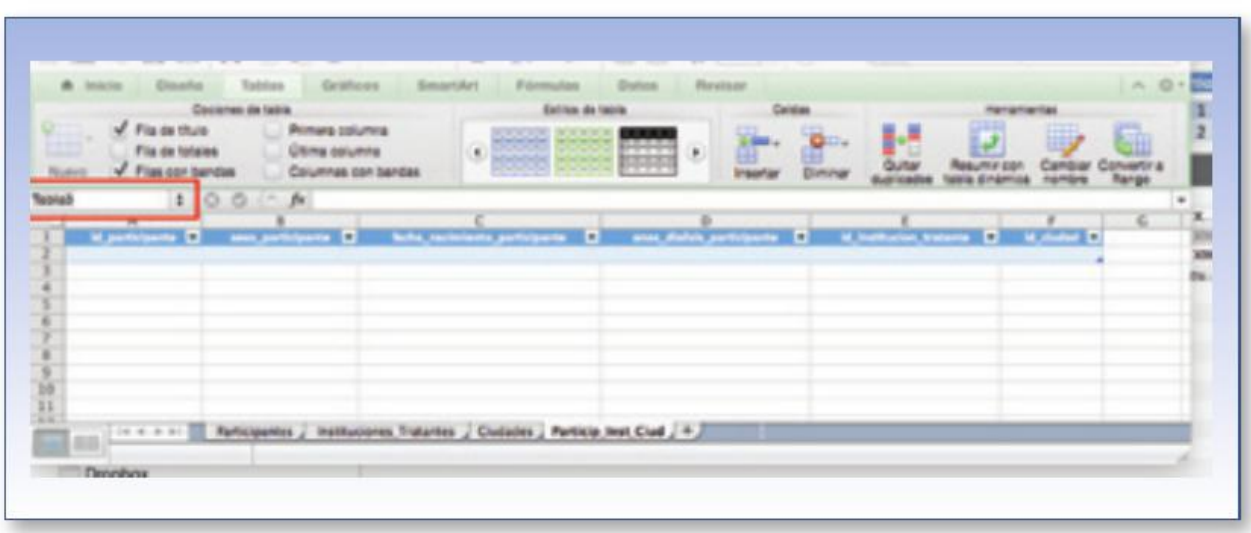


Figura 32.9 Cambio del nombre de una tabla en Excel para Mac 2011.

CÓMO VINCULAR LOS VALORES

Para vincular los valores provenientes de la tabla origen de los registros (en el caso del ejemplo, la tabla ***Participantes***), se pueden llevar a cabo los siguientes pasos:

1. Seleccione la primera celda de datos del campo que se va a vincular. En el caso del campo ***id_participante***, se trata de la celda A2.
2. Haga clic en el campo de fórmula (resaltado en la **figura 32.10**).
3. Ingrese =
4. Vaya a la hoja correspondiente a la tabla de origen del campo, y haga clic en la primera celda de datos del campo. En el caso del campo ***id_participante***, se trataría de la celda A2 de la hoja ***Participantes***. El resultado debe ser que Excel ingrese la fórmula: *=Participantes[@[id_participante]]*. Si la tabla ***Participantes*** contiene los registros que se muestran en la **figura 32.2**, el resultado de dicha fórmula debe ser el número *1*, correspondiente con el valor del campo *id_participante* del primer registro.

Cómo vincular valores de otras tablas

Para vincular los valores del campo ***id_institucion_tratante*** es necesario buscar dichos valores en la tabla ***Instituciones_Tratantes***. Si el lector siguió los ejemplos del capítulo anterior, la tabla *Participantes* debe contener un campo llamado *id_institucion_tratante*. Sin embargo, como lo muestra la **figura 32.11**, ese campo no solo contiene el identificador de la institución tratante, sino una concatenación de este y del nombre de la institución, correspondiente con el campo ***seleccion*** de la tabla ***Instituciones_Tratantes*** (como se muestra en la **figura 32.12**).

Para vincular el verdadero campo ***id_institucion_tratante***, de la tabla ***Instituciones_Tratantes***, se pueden adelantar los siguientes pasos:

1. Vaya a la hoja correspondiente a la tabla *Instituciones_Tratantes*, y seleccione el campo ***seleccion*** haciendo clic en el encabezado de la columna D (resaltado en la **figura 32.13**).
2. Utilice la opción *Cortar* (Ctrl + X en Windows o Cmd + X en Mac OSX).

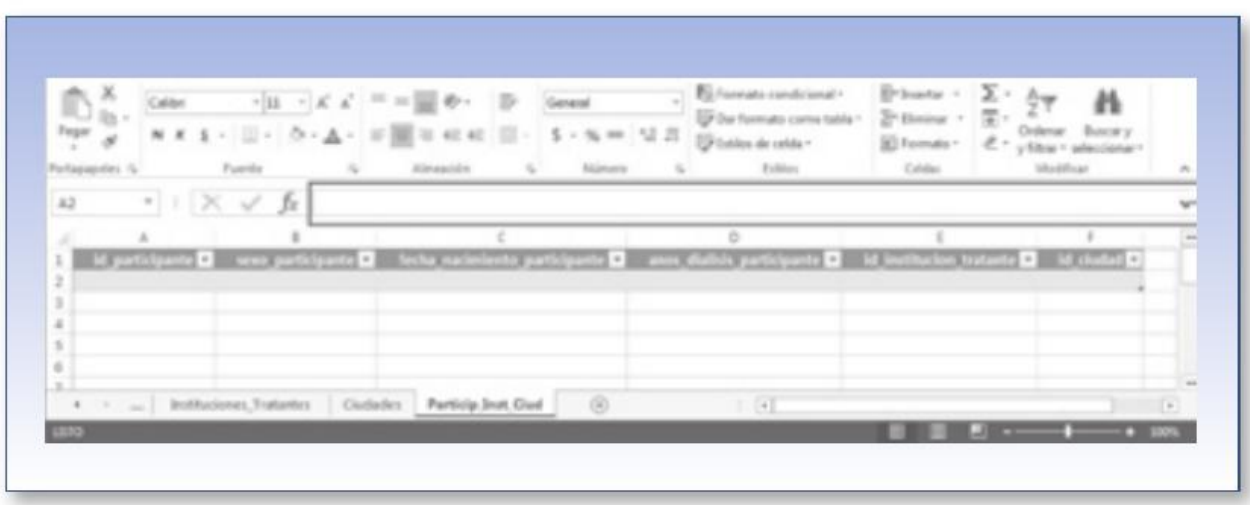


Figura 32.10 Campo de fórmula.



Figura 32.11 Tabla *Participantes*, del estudio de ejemplo.



	A	B	C	D
1	id_institucion_tratante	nombre_institucion_tratante	id_ciudad	seleccion
2	1	Hospital Universitario San Ignacio	1 - Bogotá	1 - Hospital Universitario San Ignacio
3	2	Hospital de San José	1 - Bogotá	2 - Hospital de San José
4	3	Hospital Pablo Tobón Uribe	2 - Medellín	3 - Hospital Pablo Tobón Uribe

Figura 32.12 Tabla *Instituciones_Tratantes*, del estudio de ejemplo.



	A	B	C	D
1	id_institucion_tratante	nombre_institucion_tratante	id_ciudad	seleccion
2	1	Hospital Universitario San Ignacio	1 - Bogotá	1 - Hospital Universitario San Ignacio
3	2	Hospital de San José	1 - Bogotá	2 - Hospital de San José
4	3	Hospital Pablo Tobón Uribe	2 - Medellín	3 - Hospital Pablo Tobón Uribe

Figura 32.13 Selección del campo seleccion de la tabla *Instituciones_Tratantes*.

3. Seleccione la primera columna de la hoja haciendo clic en el encabezado de la columna A.
4. Haga clic en el encabezado de la columna A con el botón secundario del ratón.
5. Utilice la opción *Insertar celdas cortadas*. Estos primeros pasos mueven el campo *seleccion* a la primera columna de la tabla. Más adelante se explicará por qué es necesario que dicho campo esté ubicado en tal posición.
   Vaya a la hoja correspondiente a la tabla *Particip_Inst_Ciud*.
6. Seleccione la primera celda de datos del campo *id_institucion_tratante*, e ingrese la siguiente fórmula: *=BUSCARV(Participantes[@[id_institucion_tratante]],Instituciones_Tratantes,2,FALSO)*

Si la tabla ***Instituciones_Tratantes*** contiene los registros que se muestran en la **figura 32.13**, el resultado de la fórmula debe ser el número *1*, correspondiente con el valor del campo ***id_institucion_tratante*** del primer registro.

La fórmula ingresada en el punto 7 le indica a Excel que debe buscar un registro en la tabla *Instituciones_Tratantes* cuyo valor en el primer campo corresponda con el valor del campo *id_institucion_tratante,* de la tabla ***Participantes***. Una vez encontrado el campo, Excel debe recuperar el valor del segundo campo del registro (que en nuestro caso corresponde con el campo *id_institucion_tratante*).

El proceso descrito puede sonar un poco enredado al comienzo, pero, en realidad, es bastante simple. Para el primer registro de la tabla ***Particip_Inst_Ciud***, Excel utiliza el valor "*1 - Hospital Universitario San Ignacio*" (correspondiente con el valor del campo ***id_institucion_tratante,*** del primer registro de la tabla *Participantes*) para buscar un registro en la tabla ***Instituciones_Tratantes***. Una vez encontrado el registro, Excel recupera el valor "1", almacenado en el segundo campo del registro hallado (correspondiente con el valor del campo *id_institucion_tratante* del primer registro de la tabla ***Instituciones_Tratantes***).

La función ***BUSCARV***, de Excel, requiere que el campo con el cual se hace la búsqueda se ubique en la primera columna de la tabla; esa es la razón por la cual se requieren los pasos 1-5.

A algunos lectores les puede llamar la atención el último parámetro de la función BUSCARV. El valor de este último parámetro le indica a Excel si debe hacer una búsqueda aproximada. Al fijar este parámetro en FALSO, en el paso 7, se le está indicando a Excel que debe buscar un registro cuyo valor del primer campo corresponda exactamente con el valor proporcionado en el primer parámetro.

La función *BUSCARV* se puede usar recursivamente. Gracias a ello, se puede vincular el valor del campo *id_ciudad*, aun cuando la tabla *Ciudades* solo se relaciona indirectamente con la tabla *Participantes* (como se muestra en el diagrama de la **figura 32.14**). Para hacer eso, basta con repetir los pasos 1-6 y reemplazar la fórmula del paso 7:

=BUSCARV(BUSCARV(Participantes[@[id_institucion_tratante]],Instituciones_Tratantes,4,FALSO),Ciudades,2, FALSO)

La parte interna de la fórmula (en azul) recupera el valor del campo *id_ciudad* de la tabla *Instituciones_Tratantes* (correspondiente con el campo *seleccion* de la tabla *Ciudades*) con base en el campo *id_institucion_tratante* de la tabla *Participantes*. Una vez recuperado dicho valor, la parte externa de la fórmula (en rojo) lo utiliza para recuperar el identificador de la ciudad.

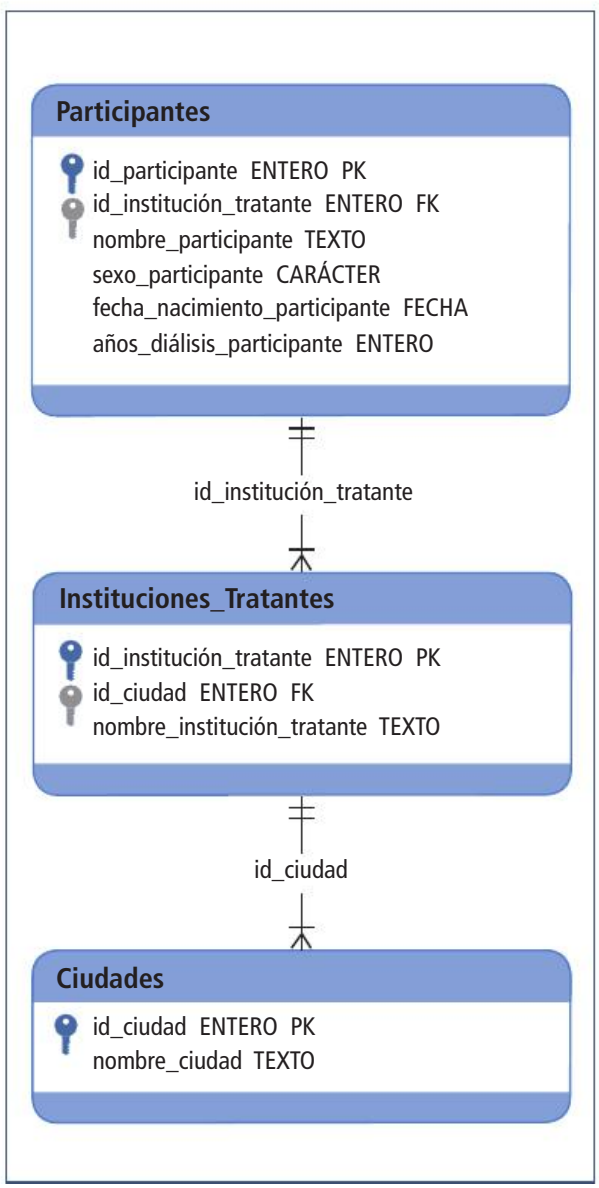


Figura 32.14 Diagrama de entidad-relación que representa a los participantes, las instituciones tratantes y las ciudades del estudio de ejemplo.

Transformación y vinculación de campos del tipo fecha

Los campos que almacenan fechas requieren atención especial, pues el formato en el que se presentan las fechas en Excel no suele ser útil para análisis estadísticos. Una alternativa para solucionar tal problema es separar la fecha en sus componentes (año, mes y día).

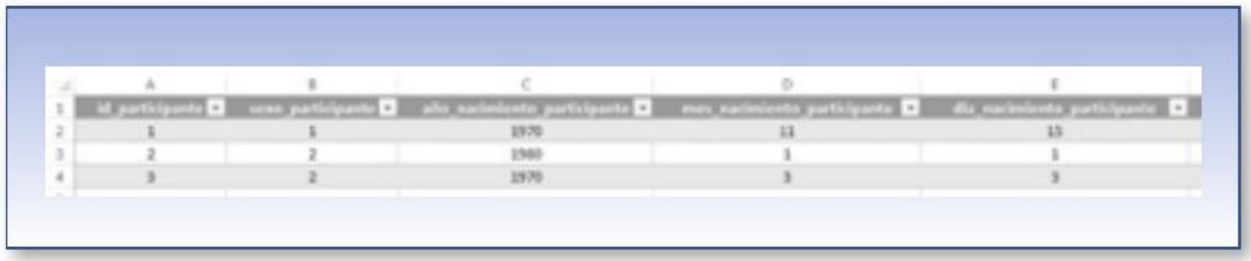
Para aplicar esta técnica al vincular el campo *fecha_nacimiento_participante*, se pueden desarrollar los siguientes pasos:

1. Vaya a la hoja correspondiente a la tabla *Particip_Inst_Ciud*.
2. Cree tres nuevos campos insertando tres columnas con encabezados: *año_nacimiento_participante*, *mes_nacimiento_participante* y *dia_nacimiento_participante*.
3. Seleccione la primera celda de datos del campo *año_nacimiento_participante*, e ingrese la siguiente fórmula: *=AÑO(Participantes[@[fecha_nacimiento_participante]])*
4. Seleccione la primera celda de datos del campo *mes_nacimiento_participante*, e ingrese la siguiente fórmula: *=MES(Participantes[@[fecha_nacimiento_participante]])*
5. Seleccione la primera celda de datos del campo *dia_nacimiento_participante*, e ingrese la siguiente fórmula: *=DIA(Participantes[@[fecha_nacimiento_participante]])*

El resultado debería verse como la tabla que se presenta en la **figura 32.15**.

Cómo vincular los datos de los demás registros

Hasta este punto, se han vinculado los valores correspondientes con el primer registro de la tabla que se va a exportar. Pero, ¿qué pasa con los valores correspondientes con los demás participantes? Para vincular los valores de los demás registros, basta con arrastrar el marcador de la tabla (resaltado en la **figura 32.15**), hasta que esta contenga tantos registros como la tabla de origen. El resultado debería verse como la tabla que se presenta en la **figura 32.16**.



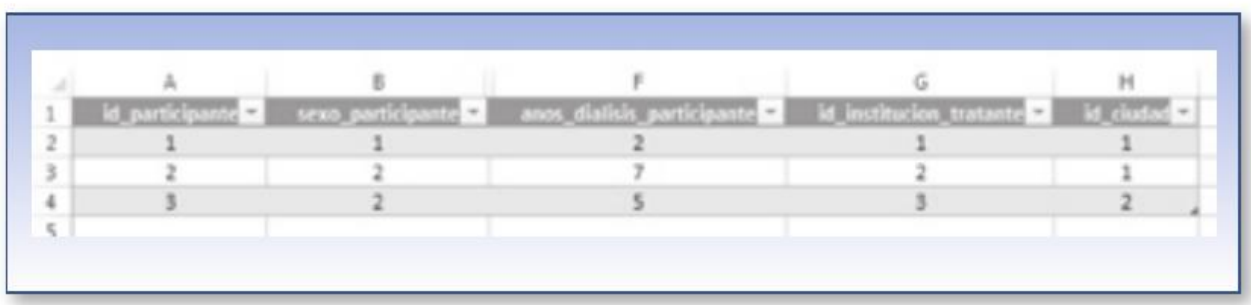
	A	B	C	D	E
1	id_participante	sexo_participante	año_nacimiento_participante	mes_nacimiento_participante	dia_nacimiento_participante
2	1	1	1970	11	15
3	2	2	1980	1	1
4	3	2	1970	3	3

Figura 32.15 Campos vinculados a la fecha de nacimiento del participante.



	A	B	F	G	H
1	id_participante	sexo_participante	anos_dialisis_participante	id_institucion_tratante	id_ciudad
2	1	1	2	1	1
3					

Figura 32.16 Vinculación de los valores para los demás registros de la tabla para exportar.*



	A	B	F	G	H
1	id_participante	sexo_participante	anos_dialisis_participante	id_institucion_tratante	id_ciudad
2	1	1	2	1	1
3	2	2	7	2	1
4	3	2	5	3	2
5					

Figura 32.17 Tabla Particip_Inst_Ciud.*

* En las tablas presentadas en las **figuras 32.16** y **32.17**, se han ocultado las columnas C y D (correspondientes a los campos *ano_nacimiento_participante, mes_nacimiento_participante y dia_nacimiento_participante*), para reducir el tamaño de las imágenes.

EXPORTACIÓN DE LOS DATOS

El formato de intercambio de datos más comúnmente usado por los paquetes estadísticos es csv. En dicho formato, cada registro se representa por una línea de texto, en la cual los valores de los campos se separan por comas. Para exportar la tabla *Particip_Inst_Ciud* en formato csv, se pueden llevar a cabo los siguientes pasos:

1. Grabe el archivo Excel utilizado la opción *Archivo* → *Guardar*.
2. Vaya a la hoja correspondiente a la tabla *Particip_Inst_Ciud*.
3. Utilice la opción Archivo → Guardar como.
4. En Excel 2013, debe ver una pantalla como la que se muestra en la **figura 32.18**. Haga clic en el botón *Examinar*. Este paso no es necesario en Excel 2010 ni en Excel para Mac 2011.
   En este punto debe obtener una ventana similar a la que se muestra en la **figura 32.19**:
5. Seleccione la carpeta en la que quiere guardar el archivo.
6. En el campo *Tipo*, seleccione *CSV (MS-DOS)*.
7. En el campo Nombre del archivo, ingrese *Particip_Inst_Ciud.csv*
8. Haga clic en el botón Guardar.
   En este punto de Excel se debe mostrar la ventana que se presenta en la **figura 32.20**; cabe advertir que el formato csv solo puede almacenar los datos de una sola hoja.
9. Haga clic en el botón *Aceptar*.

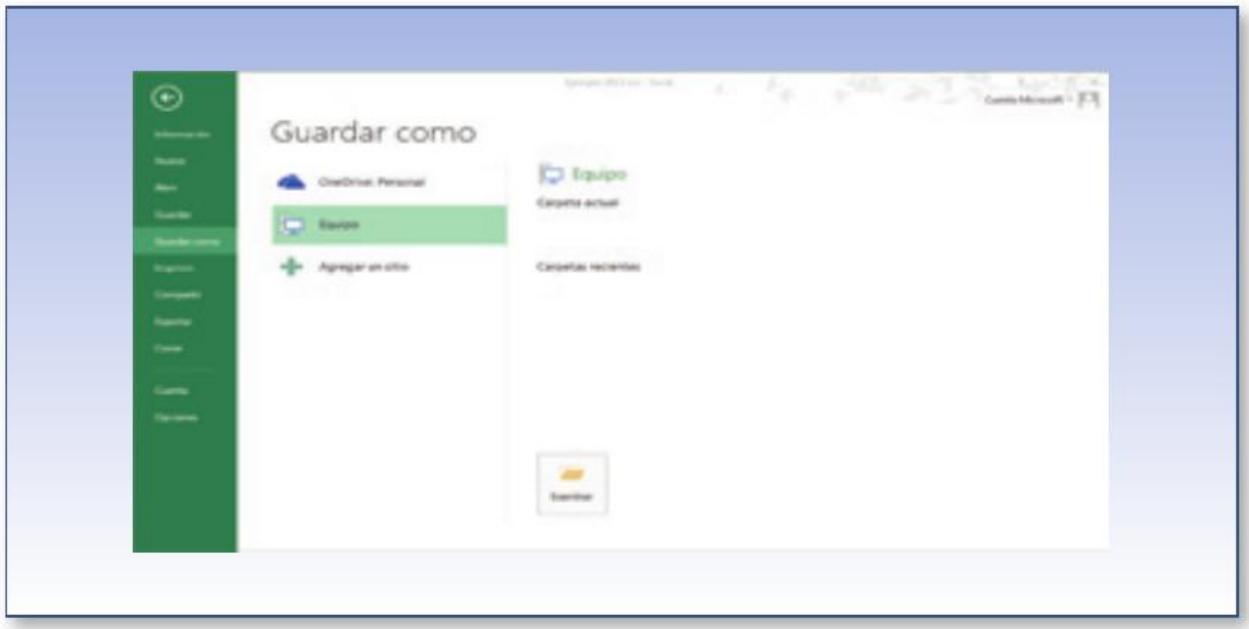


Figura 32.18 Ventana Guardar como en Excel 2013.

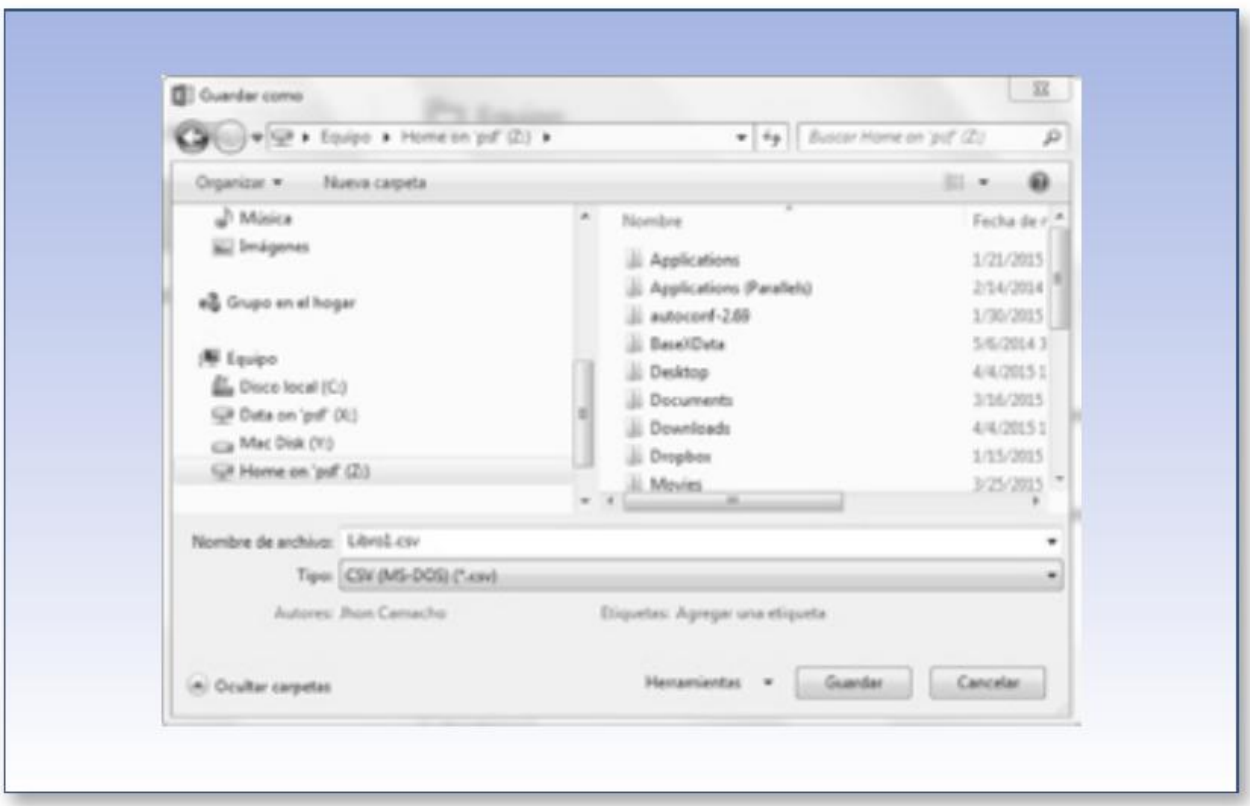


Figura 32.19 Ventana Guardar como.

En este punto de Excel debe mostrar la ventana que se presenta en la **figura 32.21**; cabe advertir que al grabar el archivo en formato **csv** se perderán algunas características del libro.

10. Haga clic en el botón *Aceptar*.

Una vez se cumple el paso 10, Excel estará trabajando con el archivo **csv, y no con el libro** que contiene la base de datos. Debido a ello, es muy importante grabar la base antes de iniciar el proceso de exportación (paso 1).

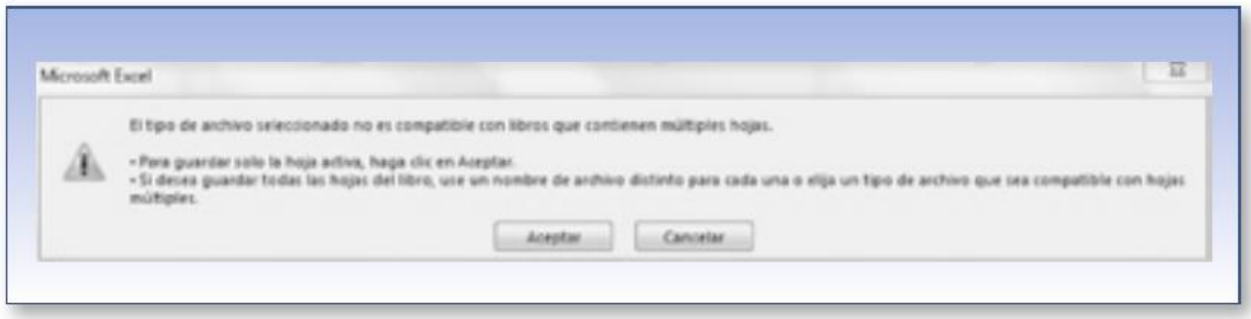


Figura 32.20 Advertencia acerca de libros con múltiples hojas al grabar en formato csv.

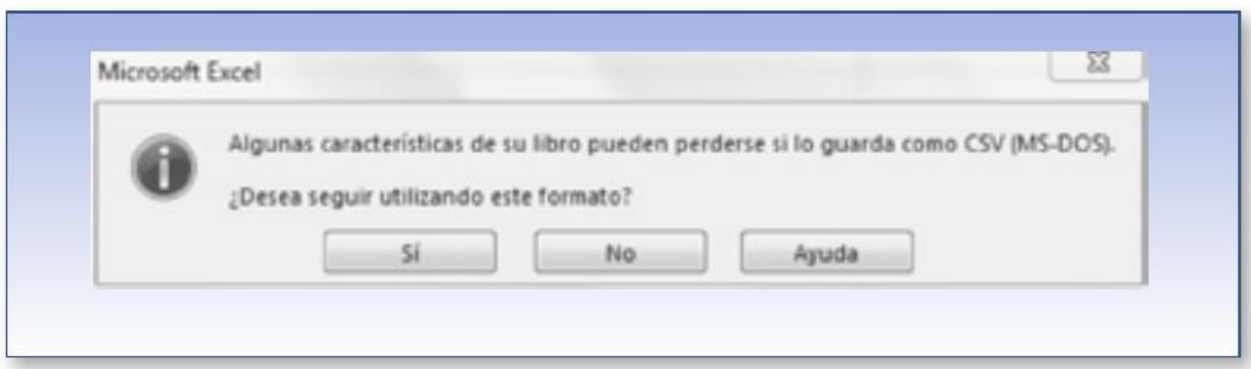


Figura 32.21 Advertencia acerca de la pérdida de características al grabar en formato csv.

CAPÍTULO

33

Presentación estadística de datos

MARTÍN ALONSO RONDÓN SEPÚLVEDA

INTRODUCCIÓN

En cualquier estudio clínico o investigativo se obtiene información respecto a una o varias cualidades o características de interés para uno o más individuos, elementos o cosas. Esta información sirve de sustento y de apoyo para la toma de decisiones. Para ello, es necesario que la información recolectada sea de buena calidad, y, a la vez, que sea analizada e interpretada de la mejor manera posible[1].

Para asegurar una buena la calidad de la información, es importante que todas y cada una de las variables recolectadas sean revisadas por medio de una inspección visual, así como de un análisis exploratorio de los datos. Con este proceso se busca identificar posibles errores o inconsistencias.

Adicionalmente, hacer una buena síntesis o resumen de la información requiere un proceso de análisis cuidadoso, y que los resultados que se presenten sean concisos, que estén los que deben estar, que se evite omitir datos importantes, que no se presenten los datos irrelevantes y que la información le permita al investigador analizar e interpretar los resultados, y al usuario final, interpretarlos y traducirlos a la práctica clínica.

El objetivo primario de la presentación de datos de un estudio es hacer que los datos "hablen por sí mismos", para facilitar la comprensión, la interpretación y la identificación de hallazgos o de relaciones que habitualmente es imposible lograr partiendo de los datos crudos. Para ello, se debe tener claridad acerca de las estrategias que se deben utilizar para resumir y presentar de manera adecuada la información. Una de ellas es mediante un análisis descriptivo de los datos, dentro de los cuales, a su vez, las tablas y las figuras son las estrategias más utilizadas.

Debido a la gran variedad de tablas y de gráficos que existe en la actualidad, surge una de las preguntas que más inquietan a clínicos e investigadores: ¿Cuál es la mejor forma de resumir, presentar e interpretar la información? No hay una respuesta única para dicha inquietud, ya que esta depende siempre de múltiples factores, tales como la pregunta de investigación, el tipo de datos y el número de variables, entre otros. Pero una guía sencilla puede apoyar muchísimo al investigador a la hora de elegir estrategias para la presentación de datos.

El objetivo principal del presente capítulo es dar una serie de pautas y recomendaciones para llevar a cabo un análisis descriptivo de la información, indicando cómo resumir y presentar información a través de tablas y figuras, y considerando para ello el número y el tipo de variables, así como su correspondiente escala de medición.

Para facilitar la numeración y la mención de los elementos gráficos, es conveniente que en un mismo artículo o en un mismo libro todas las imágenes, gráficas, esquemas, diagramas, fotos, mapas, etc. queden agrupadas bajo la categoría "figura".

Como resumir la información

Como ya se mencionó, el primer paso antes de empezar a resumir la información es inspeccionar los datos, para identificar posibles problemas de digitación, así como la posible presencia de valores atípicos o de inconsistencias. En el primer análisis de una base de datos, no es raro hallar códigos numéricos que representan opciones de respuesta distintas de las establecidas originalmente (por ejemplo, un valor de 3 en una variable que ha sido codificada solamente como 1 o 2), o valores clínicamente imposibles (como tener un valor de 100 g en peso al nacer).

El primer paso antes de intentar cualquier análisis es la limpieza de la base de datos, y un análisis exploratorio ayuda muchísimo. Al evaluar promedios, valores máximos y mínimos de una variable, podrán identificarse errores (por ejemplo: peso en un adulto de 7,5 kg; presión sistólica de 750 mmHg; potasio sérico de 73,2 mg/dl) o inconsistencias (un hombre con menopausia, o con historia de resección de ovarios). Este análisis es fundamental, ya que cualquier error, inconsistencia o valor extremo (peso en un adulto de 163,3 kg, o altura de 1,12 m) deben ser revisados, corregidos o confirmados antes de cualquier análisis. Una vez hecha tal revisión y tomados los correctivos necesarios, puede procederse al análisis estadístico[1,2]. Y los resultados de los análisis, que son el motivo de este capítulo, requieren ser presentados para su análisis o su uso final.



En general, las mencionadas medidas se pueden presentar recurriendo a tres estrategias: 1) de manera textual, 2) a través de tablas y 3) de manera gráfica[3].

Habitualmente, se recurre a una combinación de las tres estrategias, pero debe tenerse en cuenta que hay algunas reglas básicas: debe buscarse la estrategia más apropiada para cada tipo de comunicación; se deben evitar las repeticiones innecesarias (explicación completa en texto, con la misma información en una tabla, por ejemplo); debe presentarse la información positiva o negativa mientras sea relevante, y evitar, en lo posible, presentar datos que, no obstante ser verdaderos y provenientes del estudio, no aporten al análisis, a la interpretación o a la generación de conclusiones ni a la comprensión general del proyecto; finalmente, debe buscarse siempre la máxima simplicidad, sin sacrificar la calidad de la información. La presentación debe ser suficiente para permitir un enfoque apropiado de los resultados, además de correcta y comprensible[2].

Para explicar cada una de las estrategias, se utilizará un ejemplo. Los datos que se presentan a continuación **(tabla 33.1)** corresponden a la información de una muestra aleatoria de 30 niños recién nacidos en un determinado hospital en 2001, y de quienes se registró información sobre las siguientes variables: sexo, edad gestacional (en semanas), peso al nacer (en gramos), estancia hospitalaria (en días), número de hermanos y estado de salud al nacer.

Al evaluar la **tabla 33.1**, se puede ver que, si bien toda la información se encuentra en formato numérico, esta corresponde a variables tanto de tipo *cualitativo* (ya sea nominal u ordinal) como de tipo *cuantitativo* (de intervalo o de razón), y algunas de ellas son continuas, y otras, discretas. Es importante hacer esta aclaración, pues dependiendo del tipo de variable, así mismo se exploraran cuál o cuáles son las mejores estrategias para presentar y resumir dicha información. Enseguida se presentan en detalle las principales estrategias que pueden utilizarse para resumir la información del caso[1,4,5].

Análisis de manera textual



Esta es, quizá, la forma más fácil y sencilla, pero no por eso siempre la más adecuada, para presentar y resumir información.

Su fortaleza está en que toda la información se presenta de forma narrativa, y lo único que debe hacer el destinatario es leer para comprender los resultados.

No obstante lo anterior, y aunque suene sencillo, presentar un análisis narrativo tiene grandes retos. Por ejemplo, tratar de contar en

Tabla 33.1 Información de 30 niños recién nacidos en un determinado hospital en 2014

Id	Sexo	Edad gestacional (en semanas)	Peso al nacer (en gramos)	Número de hermanos	Estado de salud al nacer	Número de pulsaciones por minuto	Estancia hospitalaria (en días)
1	2	28	1214	0	2	72	54
2	2	40	3400	0	2	82	2
3	1	37	2100	1	1	142	6
4	2	40	1820	2	2	178	16
5	1	39	2288	1	1	176	11
6	1	39	3400	4	1	185	1
7	1	40	3100	1	1	116	1
8	2	39	3400	0	1	119	4
9	2	40	3500	3	1	174	2
10	2	30	1600	2	2	115	6
11	2	38	2600	2	1	169	6
12	2	37	1100	1	3	175	27
13	1	35	1820	1	1	100	79
14	2	33	1806	1	1	90	19
15	2	39	3282	1	1	115	2
16	2	37	2400	1	1	135	1
17	1	41	2920	1	1	104	5
18	2	39	2350	0	1	111	3
19	2	37	3240	0	1	136	2
20	1	38	2960	1	1	113	4
21	1	29	1060	0	3	71	47
22	2	39	3360	0	1	113	4
23	2	34	1460	1	2	85	9
24	2	38	1960	1	1	126	8
25	1	38	2700	1	1	121	2
26	2	38	2320	1	1	106	11
27	2	38	2175	1	1	187	8
28	1	30	1170	1	3	75	2
29	2	39	2687	0	1	189	19
30	1	35	2329	0	1	101	2

Id: identificador del niño. Sexo: 1. Masculino 2. Femenino. Estado: Estado de salud al nacer: 1. Bueno 2. Regular 3. Malo.

palabras lo que sucedió en el estudio implica que el investigador debe tener facilidad para la escritura, y, además, tenerla para que su descripción sea lógica y comprensible. Por otra parte, a medida que el número de variables de interés se incrementa, la tarea se va haciendo más dispendiosa, hasta llegar, incluso, a no ser viable; en especial, por el considerable consumo de tiempo que ello implica.

Como ejemplo de un análisis textual, al describir la información hallada en las variables cualitativas (que en nuestro ejemplo corresponden a las variables sexo y estado de salud al nacer), podemos decir que la muestra en estudio se halla compuesta por 30 recién nacidos, de los cuales 11 son de sexo masculino, y 19, de sexo femenino. Respecto al estado de salud al nacer, se encontró que en 22 de ellos era bueno; en 5,

regular, y en 3, malo. El problema de dar dichos valores absolutos es que muchas veces no se tiene claridad sobre cuánta diferencia hubo entre los dos grupos, por lo cual muchas veces se recomienda presentar esa información a través de medidas de resumen, como, por ejemplo, porcentajes (o promedios, en el caso de las variables cuantitativas) que permitan tener una estimación más específica de los resultados, de tal manera que sean más comprensibles. Dicha estrategia es ideal, especialmente, cuando los datos son de tipo cuantitativo, o cuando el volumen de información es grande.

Pese a lo mencionado, como siempre se debe recurrir al cálculo de medidas de frecuencia o de resumen, surge la inquietud de si no sería mejor hacer esa descripción a través de tablas o de gráficos. Lo ideal es encontrar el balance perfecto que nos permita resumir la información de la manera más eficiente utilizando para ello una combinación de las tres estrategias.

Tablas y gráficos

Las tablas y los gráficos son dos métodos utilizados para resumir y presentar información de la manera más clara, precisa y organizada posible. Las tablas se caracterizan, principalmente, por describir la información de una manera esquemática, mientras los gráficos buscan presentar la misma información de una manera más visual.

Es común que los gráficos utilicen la información contenida en las tablas para su creación; por ello, es importante considerar dicho aspecto para no duplicar la información al momento de reportar los resultados. A continuación se describirán en términos más formales ambos métodos.

Tablas de frecuencias

El objetivo principal al construir tablas de frecuencia es sintetizar de manera lógica y coherente la información. Se busca hacer un arreglo ordenado de los datos en filas y columnas, de la manera más sencilla posible, para que sea fácil y rápidamente comprensible.

Antes de hacer una tabla, es necesario evaluar si su construcción tiene sentido o no, pues muchas veces, al revisar los datos pueden encontrarse cifras de poca importancia, o de escasa o ninguna variación, y que, probablemente, podrían describirse mejor con un texto, o bien, que pueden no necesitar ser informadas por separado, y tal vez sea mejor reagruparlas. A continuación se darán los pasos generales para crear tablas.

Partes de una tabla

Para construir una tabla, es necesario considerar algunos aspectos importantes que permitan su lectura y su comprensión. Enseguida se mencionan los atributos más relevantes de una tabla:

- Debe llevar un título: **Es importante que el título** sea claro, conciso y preciso, y, a la vez lo más completo posible. Esto significa que el título debe contener información detallada sobre qué, cómo, dónde y cuándo fue tomada la información que se está presentando. Por ejemplo, en el título de la **tabla 33.1**, **el "qué"** corresponde a la ***distribución de frecuencias*** *de 40 recién nacidos*; **el "cómo", a por sexo;** el "dónde", a en un hospital, y el "cuándo", a en 2001.
- Descripción de la información que se presenta en las filas de la tabla: **Es importante aclararle al** lector, qué información específica se le está mostrando, por cuanto es posible que muchas variables tengan las mismas opciones de respuesta; por lo tanto, se debe especificar a cuál de ellas corresponde (por ejemplo, 1 en sexo puede significar sexo masculino, pero en diabetes ***mellitus*** el valor 1 puede significar presencia de la enfermedad). Generalmente, la primera fila corresponde a los nombres de cada una de las columnas que se van a presentar.
- Descripción de la información que se presenta en las columnas de la tabla: **Es así mismo importante** detallar con exactitud qué información se presenta en las columnas, pues en el caso de una sola variable, esta puede ser la frecuencia absoluta o el porcentaje (número de recién nacidos con bajo peso, o el porcentaje del total), mientras en el caso de dos variables puede representar información más compleja (por ejemplo, el número de diabéticos en el grupo de personas con infarto).
- Cuerpo de la tabla: **Es la información numéri**ca reportada que se especifica en el título; es decir, la información de interés. Dicha

información se puede dar como frecuencias absolutas, frecuencias relativas, porcentajes, razones o tasas.

- Notas de pie de tabla: En muchas ocasiones, la descripción que se hace de las filas, las columnas o el cuerpo de la tabla es insuficiente, debido al poco espacio con el cual se cuenta. Es allí donde aparecen figuras o símbolos que indican que al final de la tabla se hará una descripción más detallada o una aclaración, las cuales se conocen como notas al pie de tabla. Un ejemplo se presenta en la **tabla 33.1**.
- Fuente: No es raro que en la mayoría de las investigaciones el autor reporte información extraída de otras fuentes o de otros estudios. En esos casos, es importante que al crear la tabla quede registrado que la información no es original y proviene de una fuente secundaria. Por ejemplo: ***Fuente: DANE-Censo*** *Nacional de Población y Vivienda 2003*. Es importante anotar que cuando la información proviene del estudio que está siendo informado no se requiere aclarar la fuente.

Aunque hay una gran variedad de tipos de tablas, y estas dependen de múltiples factores como el tipo y el número de variables en análisis, así como del propósito del estudio (descripción, asociación, etc.), en general, estas buscan resumir la información de dos aspectos principales: contar el número de veces que se presenta cada una de las categorías de un evento de interés (el cual, a su vez, se conoce como frecuencia absoluta) o la proporción de eventos que se presentaron en cada una de esas categorías (conocido como *frecuencia relativa*).

Esta última muchas veces se presenta como *frecuencia porcentual* (que corresponde a las frecuencias relativas multiplicadas por 100).

A partir de las frecuencias mencionadas en el párrafo anterior, se pueden generar sus correspondientes frecuencias acumuladas, lo que corresponde a la suma de las frecuencias (absolutas, relativas o porcentuales) de los valores correspondientes a las categorías menores o iguales a la que se está considerando. En la siguiente sección se hará una descripción más detallada de cada una de dichas frecuencias, considerando, además, el tipo y el número de variables.

Gráficos

Los gráficos son representaciones visuales, que buscan, a través de estrategias expresivas, permitirle al lector, con una mirada, identificar y comprender las características más importantes de los resultados (como dijo Barnard [1927], "una imagen vale más que mil palabras").

Para su construcción, es necesario identificar claramente cuál es su propósito, pues existe una amplia variedad de posibilidades al respecto, y siempre es necesario tener claro lo que se pretende mostrar y cómo hacerlo de la manera más sencilla posible.

En general, los gráficos se utilizan para evaluar la forma de la distribución de las variables, así como para hacer comparaciones y evaluar tendencias.

Partes de un gráfico

Al igual que para las tablas, al momento de construir gráficos se deben tener en cuenta las siguientes consideraciones:

- Título: Debe cumplir las mismas condiciones presentadas para las tablas.
- Descripción de los ejes: Se debe especificar la información que se presenta, por lo cual es de mucha utilidad incluir un título que identifique claramente a qué variable corresponde cada uno de los ejes del gráfico, así como las unidades en las cuales se están presentando (por ejemplo, frecuencias absolutas, porcentajes, etc.).
- Área del gráfico: Es la forma como se representa la información. Siempre depende del tipo de gráfico que se esté haciendo.
- Notas a pie de página: Cumplen las mismas características presentadas para las tablas.
- Leyenda: En caso de ser necesario, se debe incluir un recuadro que permita identificar dentro del área del gráfico cuáles son las categorías de análisis.
- Líneas de división: También se deben incluir en caso de que el tipo de gráfico lo permita y sea adecuado. El detalle al cual se llegue en las líneas depende de la información que se presenta, de los valores máximo y mínimo y de las variaciones existentes. Deben evitarse tanto el exceso como la pobreza innecesaria.

- Fuente: Cumple las mismas características presentadas anteriormente para las tablas.

Al igual que en el caso de las tablas, la selección del mejor gráfico depende de múltiples factores, como el tipo de variable (cualitativa o cuantitativa), el número de variables que se analizarán (univariado, bivariado o multivariado) y el objetivo de interés de la investigación (descripción, asociación, comparación). En este escenario, en las siguientes secciones se van a presentar los principales tipos de tablas y de gráficos utilizados para resumir información considerando estos múltiples factores; inicialmente, para variables cualitativas, y posteriormente, para variables cuantitativas.

TABLAS Y GRÁFICOS PARA UNA VARIABLE CUALITATIVA

Inicialmente se explica cómo construir una tabla de frecuencias en el caso de tener una variable cualitativa, y a continuación se evaluarán algunos de los gráficos más utilizados para presentar este tipo de información.

Tabla de frecuencias para una variable cualitativa

Lo primero que debe identificarse es si la variable es de tipo nominal u ordinal, ya que en este último caso las categorías se deben poner en el orden predeterminado. Por ejemplo, para la variable "sexo", cuyas opciones de respuesta son *masculino* y *femenino*, al tratarse de una variable cualitativa nominal, no hay un orden predeterminado, y pueden ponerse según el criterio del investigador, mientras para el caso de la variable "estado de salud al nacer", sus categorías deben incluirse en el orden predeterminado o lógico (***bueno, regular, malo***).

Tanto para las variables nominales como para las ordinales, la información que por lo general se presenta corresponde a la distribución de las frecuencias absolutas, de las frecuencias relativas (o porcentajes) y de las frecuencias acumuladas de las distribuciones, ya mencionadas.

Como ejemplo de lo anterior, se va a utilizar la variable estado de salud al nacer, de la **tabla 33.1**. El primer paso para construir la tabla es hacer un bosquejo, donde se pone en las filas correspondientes a la primera columna la variable con sus respectivas categorías de respuesta. Posteriormente, en la columna dos se incluye la información correspondiente a las frecuencias absolutas; en la tercera, el porcentaje, y en las columnas cuatro y cinco, las concernientes a las frecuencias absolutas acumuladas y al porcentaje acumulado.

Una vez creado el mencionado bosquejo, se procede a completar la tabla. Para ello, en el caso de las frecuencias absolutas, se cuenta el número de veces que se presenta cada una de las opciones de respuesta. Si se recuerda el análisis hecho de manera textual, en la muestra de 30 recién nacidos, 22 se hallaban en buenas condiciones; 5, en regulares condiciones, y 3, en malas condiciones. Esta información se ubica, por lo tanto, en las casillas correspondientes de la tabla.

Posteriormente deben calcularse las frecuencias relativas, que corresponden a una proporción, la cual se define como el número de unidades de observación (recién nacidos) con una característica de interés (presentar un estado de salud bueno para el primer caso, regular para el segundo y malo para el tercero), dividida por el total de unidades de observación (total de la muestra; en este caso, 30). Por tanto, la proporción de recién nacidos en buen estado de salud en la muestra corresponde al cociente $22 / 30 = 0{,}73$; en regular estado, a $5 / 30 = 0{,}17$, y en mal estado, a $3 / 30 = 0{,}1$. En general, se recomienda, para una mejor interpretación de los resultados, presentarlos como porcentajes, lo cual equivale a multiplicar cada una de las frecuencias relativas por 100; es decir, puede afirmarse que el 73 % de los recién nacidos presentaron un buen estado de salud al nacer; el 17 %, regular estado, y el 10 % restante, mal estado. Tal información también se ubica en las casillas correspondientes de la tabla.

A continuación, se calculan las frecuencias y los porcentajes acumulados, que, como ya se mencionó, corresponden a la suma de las frecuencias de los valores correspondientes a las categorías menores o iguales a la que se está considerando. En este sentido, la frecuencia absoluta acumulada de la primera categoría, la cual se refiere a un estado de salud bueno, y dado que no existe una categoría anterior, simplemente corresponde al total de recién nacidos con esta característica en la muestra (es decir, 22, que aparece en la columna de la frecuencia absoluta).

La diferencia va a estar en la siguiente categoría, que corresponde a un estado regular, y en la que, al evaluar las categorías menores o iguales a esta, se deben sumar las frecuencias absolutas de la categoría bueno y regular (recuérdese que se trata de una frecuencia acumulada), equivalente a la suma de 22 y 5; o sea, 27 recién nacidos en la muestra presentaron un estado de saludo entre regular y bueno.

Finalmente, para la tercera categoría, al evaluar las categorías menores o iguales a esta, el resultado corresponde a todo el espectro, ya que incluye dentro de la categoría de malo las categorías de regular y bueno; es decir, es el tamaño de muestra total (n = 30). El mismo proceso se hace para el caso del porcentaje acumulado. Estos datos también son ubicados en la tabla, y como proceso final se incluye el título y se evalúa si es necesario agregar notas al pie de la tabla o fuentes. Una vez cumplido dicho paso, está listo el producto final, que se presenta en la **tabla 33.2**.

Gráficos para una variable cualitativa

Una vez construida la tabla, el investigador debe evaluar si esta es lo suficientemente clara o si valdría la pena presentar dicha información de una manera más visual, por lo que se pueden considerar algunos gráficos para su presentación.

Los gráficos más utilizados para presentar la información de variables cualitativas son: 1) gráficos de barras simples, 2) pictogramas y 3) gráficos de torta o *pie*[2,6,7].

Gráfico de barras simples

Este gráfico es uno de los más utilizados, ya que, además de ser fácil de construir, también permite presentar la información de distribuciones de frecuencias para variables cualitativas tanto nominales como ordinales, al igual que para variables cuantitativas discretas; en algunas ocasiones, incluso, se utiliza para la representación de series cronológicas o históricas.

Se caracteriza porque consta de barras (rectángulos), que identifican cada una de las categorías de respuesta, las cuales, por tratarse de valores discretos, siempre deben estar separadas entre sí. Estas barras siempre deben tener la misma anchura, y pueden ser presentadas de manera horizontal o vertical. La altura de cada barra debe ser proporcional a la frecuencia presentada para cada una de las categorías. No son recomendables para mirar tendencias en variables cualitativas nominales.

Para su construcción, a diferencia de las tablas, es necesario decidir específicamente qué tipo de información es el que se presentará: si porcentajes, o frecuencias absolutas, o frecuencias relativas. Como ejemplo, en la **figura 33.1** se presenta un gráfico de barras simples con la información correspondiente a las frecuencias absolutas que se encuentran en la **tabla 33.2**.

Pictograma

Un pictograma es un gráfico que utiliza símbolos, imágenes o figuras alusivas a un tema que se quiere representar, y del cual se busca que sea llamativo para el lector.

El objetivo del pictograma es presentar información relevante que, se espera, pueda ser captada con rapidez. Para su construcción, se usa, generalmente, el gráfico de barras simple, pero se cambian los rectángulos por imágenes relativas al tema de interés; sin embargo, en ocasiones también es utilizado en otros tipos de gráficos **(figura 33.2)**.

Tabla 33.2 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del estado de salud de 30 recién nacidos en un hospital durante 2014

Estado de salud al nacer	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
Bueno	22	73	22	73
Regular	5	17	27	90
Malo	3	10	30	100

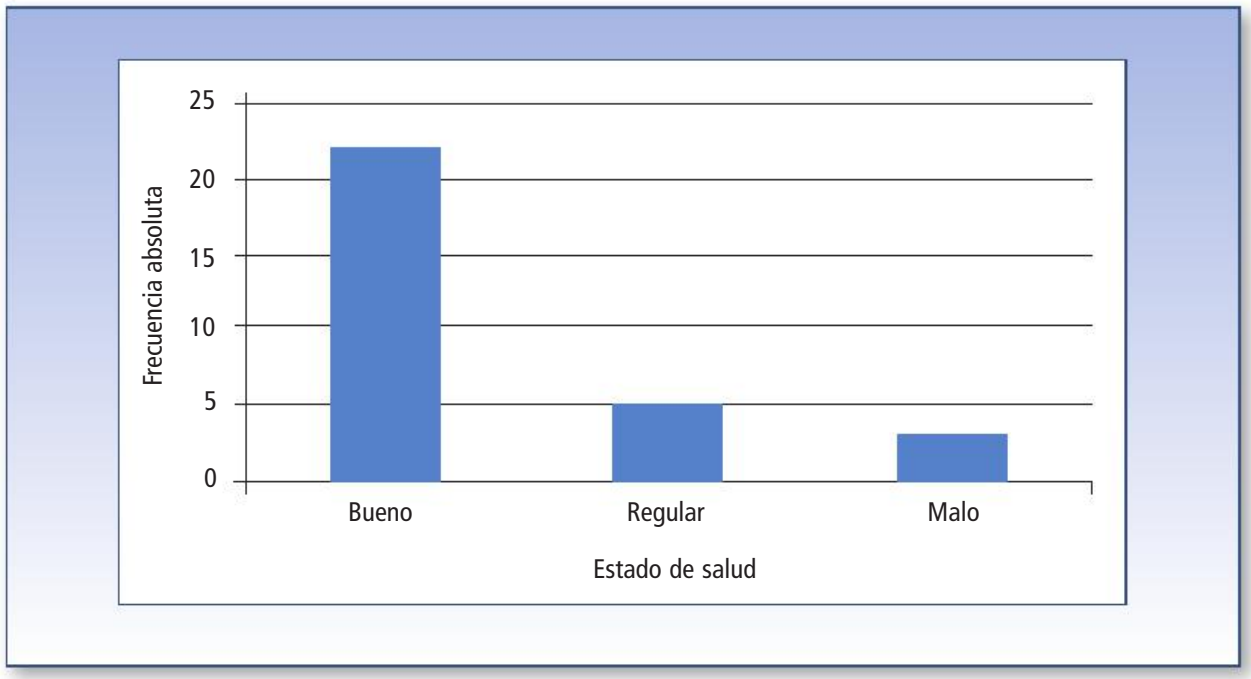


Figura 33.1 Gráfico de barras para mostrar la información "Distribución del estado de salud de 30 recién nacidos en un hospital en 2014".

Gráfico de torta, o *pie*

También conocido como ***gráfico de sectores***, es otro de los gráficos más utilizados por su fácil construcción y su fácil interpretación. Se usa, fundamentalmente, para representar distribuciones de frecuencias relativas (%); por lo tanto, es útil para comparar proporciones. Al igual que el gráfico de barras simples, el ***pie*** también se utiliza para variables cualitativas, tanto nominales como ordinales, o para variables cuantitativas discretas. Cabe tomar en cuenta una contraindicación muy importante: no debe utilizarse para representar categorías que no sean mutuamente excluyentes. No sería apropiado, por ejemplo, con el fin de presentar la frecuencia de factores de riesgo para infarto del corazón, pues cada uno de los factores tendrá una frecuencia relativa en relación con la población total (denominador), pero un fumador puede, además, ser diabético, o un hipertenso puede tener diabetes, fumar y tener colesterol elevado.

Otra de sus desventajas es que a medida que el número de categorías se incrementa (por lo general, siete o más), o que la distribución de algunas de sus categorías es muy baja, las hace difíciles de interpretar, lo cual limita su uso.

Un error que debe evitarse a toda costa es presentar este gráfico (y el de barras, su compañero de alta frecuencia en este error) en tres dimensiones. Tal vez con el propósito de hacerlo semejante a un verdadero pastel o torta, se ha hecho popular este error, que distorsiona la información, pues agrega una dimensión inexistente (el grosor de cada tajada) y dificulta interpretarla. Casi siempre, las tajadas de la parte inferior del gráfico, donde se aprecia la tercera dimensión completa, aparecerán como más grandes que las de la parte superior, aunque las cifras no correspondan con esa percepción, producto de una ilusión visual, como puede verse en la **figura 33.3**, sobre dos representaciones de los mismos datos.

Cada una de las secciones o arcos de circunferencia representa las frecuencias relativas correspondientes a las diferentes categorías respetando el orden establecido en la tabla y partiendo de un punto dado de la circunferencia (se recomienda que sea el equivalente a las 12 en el reloj).

Para su construcción, debe aplicarse una simple regla de tres. Para ello, el círculo, en total, tiene 360°, lo que equivale al total (100 %); por lo tanto, si se busca encontrar el área correspondiente a la categoría de interés (por ejemplo, en el caso de la variable sexo, podemos hacerlo para masculino), y dado que el área es equivalente a la frecuencia relativa, se encuentra el ángulo a partir de la siguiente operación: $360 \times 0{,}37 = 133°$. De la misma forma, se encuentra el ángulo para el sexo femenino, que es: $360 \times 0{,}63 = 227°$.

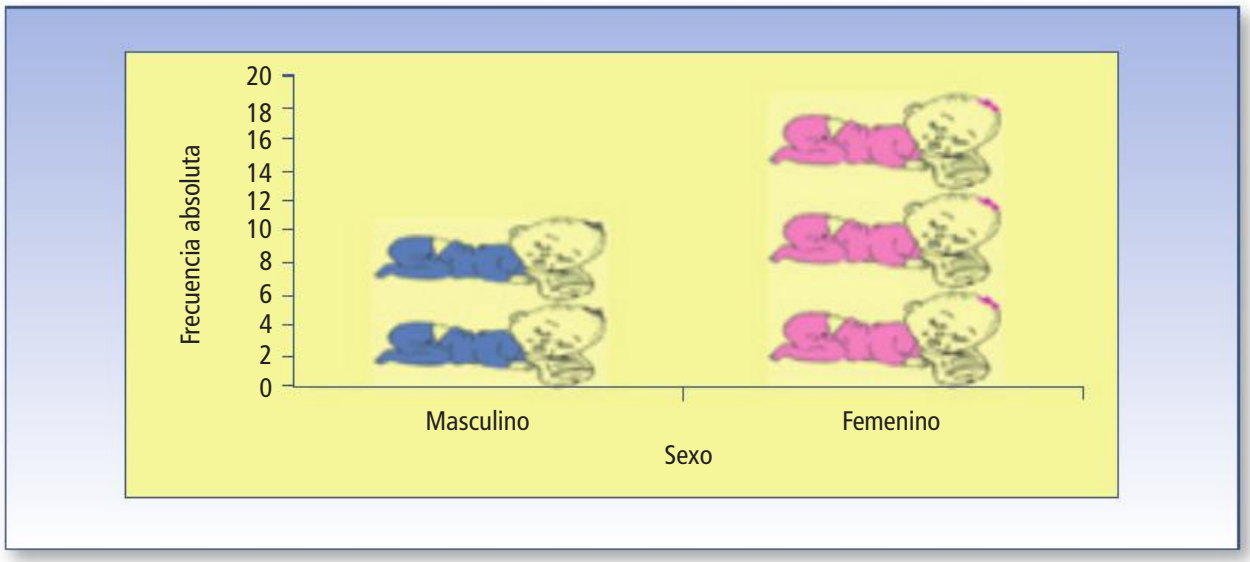


Figura 33.2 Pictograma para representar la información "Distribución del sexo de 30 recién nacidos en un hospital en 2014".

Estos ángulos se trasladan al gráfico tomando como referencia la línea que aparece a las 12, y se ubica el área correspondiente **(figura 33.4)**. La mayoría de los programas de computador hacen automáticamente los cálculos y las distribuciones de área, por lo cual el usuario debe asegurarse de que ingresa los datos completos y correctos.

TABLAS Y GRÁFICOS PARA UNA VARIABLE CUANTITATIVA

Hay una gran variedad de tablas y de gráficos para presentar visualmente variables cuantitativas, tales como las distribuciones de frecuencia, los histogramas de frecuencia, los polígonos, las ojivas, gráficos de tallo y hoja, y lo gráficos de caja y bigote, por mencionar solo algunos. Decidir cuál de ellos utilizar depende del objetivo de interés y del tipo de datos en análisis.

En el caso de las variables cuantitativas, estas se clasifican según el tipo de datos como *discretas* **(por ejemplo, número de hermanos)** o ***continuas*** **(como peso al nacer). Las variables discretas** se caracterizan porque sus valores corresponden a números enteros, mientras las continuas pueden tomar cualquier valor en un intervalo de posibilidades dado. A continuación se presenta cómo construir tablas y gráficos para variables cuantitativas, considerando para la construcción el tipo de datos.



Figura 33.3 Diferencias entre un gráfico de pastel plano y uno en tres dimensiones

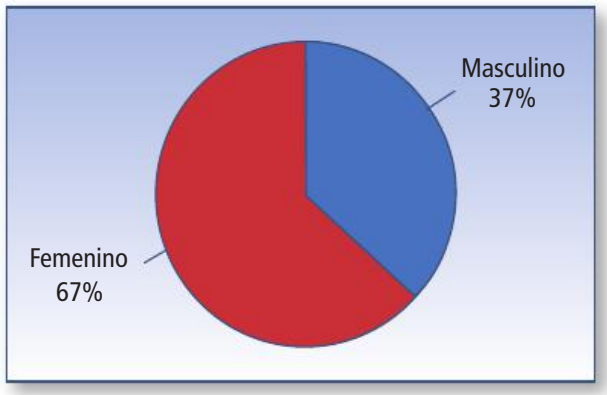


Figura 33.4 Gráfico de torta o *pie* para mostrar la información "Distribución del sexo de 30 recién nacidos en un hospital en 2014".

Tabla de frecuencias para una variable cuantitativa discreta

Al igual que para las variables cualitativas, la información que se presenta al utilizar variables discretas corresponde a la distribución de las frecuencias absolutas, de las frecuencias relativas (o porcentajes) y de las frecuencias acumuladas. En caso de presentarse pocas opciones de respuesta, se hace el mismo proceso que se siguió para variables ordinales; o sea, crear una categoría para cada opción de respuesta y calcular las frecuencias respectivas en cada una de ellas. Por ejemplo, para la variable del ejemplo (número de hermanos), al evaluar sus opciones de respuesta se ve que se pueden tomar los valores 0, 1, 2, 3 y 4. Al no ser muchas las opciones de respuesta, se puede crear una tabla que presente las cinco categorías, y se procede a llenarla. Los resultados de este proceso se presentan en la **tabla 33.3**.

Si la variable presenta muchas categorías de respuestas (por lo general, más de diez), o si la frecuencia de algunas de las categorías es muy baja o nula, se debe evaluar la posibilidad de reagrupar categorías en grupos que mantengan el orden de la variable. Por ejemplo, al analizar los datos de la tabla 3, se nota que algunas de las categorías presentan baja frecuencia (en especial, las correspondientes a 3 y 4), ante lo cual se puede plantear reconstruir la tabla reagrupando categorías. De esta manera, se podría pensar en dejar solo una, que represente a dos o más hermanos. La **tabla 33.4** presenta tal redistribución.

La misma estrategia se puede utilizar en caso de presentarse un número finito de posibilidades, como, por ejemplo, en el caso de la variable número de pulsaciones por minuto. Sin embargo, en caso de que el investigador no conozca el número de grupos que debe crear, se utilizan estrategias o reglas para determinar el número y la amplitud de las categorías.

Una de las más conocidas es la regla de Sturges[8], que le indica al investigador el número de categorías sugerido para su tabla según la cantidad de datos que quiera resumir, en el supuesto de que todas las categorías presentan la misma amplitud. La recomendación que hace Sturges es estimar el número de categorías (que se representan como k) a través de la siguiente ecuación:

$$k = 1 + 3.322 \times log_{10} n$$

Una vez obtenido el valor de k, se procede a calcular la amplitud del intervalo (A). Para ello, se calcula el rango de los valores (R), el cual se define como la diferencia entre el valor máximo y el valor mínimo observados, y dividido sobre el número de categorías (k); A=R/k.

Posteriormente, Lane sugiere una modificación a la regla de Sturges, y, al igual que la anterior, solo basa su cálculo en el tamaño de la muestra, con la siguiente fórmula.

$$k = 2 \times \sqrt[3]{n}$$

Tabla 33.3 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del número de hermanos de los 30 recién nacidos en un hospital en 2014

Número de hermanos	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
0	9	30,0	9	30,0
1	16	53,3	25	83,3
2	3	10,0	28	93,3
3	1	3,3	29	96,6
4	1	3,3	30	100,0

Tabla 33.4 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del número de hermanos de los 30 recién nacidos en un hospital en 2014 reagrupando categorías

Número de hermanos	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
0	9	30,0	9	30,0
1	16	53,3	25	83,3
2 o más	5	16,7	30	100,0

No obstante lo anterior, ambas reglas fueron muy criticadas, por no considerar aspectos estadísticos para su construcción, pues si bien asumen que la variable sigue una distribución normal, solo consideran en su cálculo el tamaño de la muestra. Ante ello, algunos autores, como Scott, Doane, Freedman y Diaconis[9,10], entre otros, han presentado reglas alternativas para su cálculo.

La alternativa que ofrece Scott es respecto a la amplitud del intervalo, pero también asumiendo la normalidad de la variable y que todos los intervalos son de igual amplitud: para su cálculo, toma en cuenta la desviación estándar y el tamaño de la muestra. El número de categorías, por lo tanto, se calcula como:

$$k = \frac{R \times (\sqrt[3]{n})}{3{,}5\,\sigma}$$

Donde R es el rango de los valores, y **σ, la** desviación estándar.

Por otra parte, Doane propone modificar la regla de Sturges, pero considerando que la variable no sigue una distribución normal. Por ello, considera para el cálculo el coeficiente de asimetría de la variable:

$$k = 1 + (3.322 \times log_{10} n) + \left\{ 3.322 \times log_{10} \left[1 + \left(\frac{|g_1(x)| \times \sqrt{(n+1)\times(n+3)}}{6\times(n-2)} \right) \right] \right\}$$

Donde $g_1(x)$ **es el coeficiente de asimetría** de la variable x.

Freedman y Diaconis, al igual que en el caso anterior, tampoco asumen la normalidad de la variable, y para hacer la estimación de la amplitud del intervalo consideran el rango intercuartílico:

$$k = \frac{R}{2 \times \left(\frac{RIC(x)}{\sqrt[3]{n}} \right)}$$

Donde RIC(x) es el Rango Intercuartílico de la variable x, y R, el rango de los datos.

Dada la gran variabilidad posible en los datos, en general, estas corresponden al número de categorías sugerido; por lo tanto, la decisión final es del investigador. En general, se recomienda tomar el número menor hallado entre las distintas alternativas.

En el caso del ejemplo, los resultados que se encontraron para el número de pulsaciones por minuto al aplicar las estrategias de Sturges y de Lane fue 6; al aplicar la de Doane, 7, y para las reglas de Scott y Freedman-Diaconis, 3. Por lo tanto, en este caso la recomendación es crear *tres* **intervalos de igual amplitud.**

A continuación se debe calcular la amplitud del intervalo, que corresponde al rango **($R = 189 - 71 = 118$) dividido por el número de** categorías **($k = 3$); se obtiene así una amplitud** de 40. Al tratarse de una variable discreta, se debe tener cuidado al momento de crear los límites de cada una de las categorías. A manera de ejemplo, puede suponerse que se tiene un valor mínimo de 10, y que la amplitud del intervalo es de 3. Esto significa que como la variable es discreta, el intervalo debería contener tres opciones de respuesta, que serían, en este caso, 10, 11 y 12. Por lo tanto, la primera categoría de dicho intervalo estaría dado por [10-12]. En la **tabla 33.5** se presentan las categorías conformadas para la variable del ejemplo.

Gráficos para una variable cuantitativa discreta

Los gráficos recomendados para este tipo de variables son los mismos que se consideran para las variables cualitativas ordinales; es decir, los gráficos de barras simples, los gráficos de torta y los pictogramas. Adicionalmente, algunos autores también recomiendan el uso de gráficos de punto para una variable, de gráficos de caja y bigote (***box plot***) **y gráficos de caja de letras y valores**[11-15].

Tabla 33.5 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del número de pulsaciones por minuto de los 30 recién nacidos en un hospital en 2014

Número de pulsaciones	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
71-110	10	33,3	10	33,3
111-150	12	40,0	22	73,3
151-190	8	26,7	30	100,0

Gráfico de punto para una variable

Este tipo de gráfico se usa con el fin principal de evaluar la forma de distribución de la variable, así como de la dispersión de los valores. Es sencillo de construir, pues, simplemente, sobre una escala métrica que contiene el rango de valores en estudio se ubica un punto cada vez que aparece un valor específico sobre ese mismo valor, y en caso de que se repita, se van apilando dichos puntos de manera vertical. En la **figura 33.5** se presentan los resultados para la variable número de hermanos. A partir de este gráfico puede decirse que, si bien la mayoría de los recién nacidos presentaron edades entre las 37 y las 41 semanas, se trata de una variable muy heterogénea (dispersión alta), y no se puede identificar fácilmente que tipo de distribución sigue esta variable.

Gráfico de box plot *(caja y bigote)*

Es uno de los gráficos más utilizados y recomendados para evaluar la forma de distribución de la variable y la homogeneidad de esta, así como para detectar la posible presencia de posibles valores atípicos tanto para variables discretas como para variables continuas.

Como su nombre lo indica, corresponde a un gráfico que se basa en una caja rectangular, la cual es dividida en dos partes por la mediana y dos líneas que se conectan a esta caja, y que, a su vez, se conocen como ***bigotes***. Si las magnitudes o los tamaños de cada una de las divisiones de la caja y de los bigotes son similares, se puede decir que la distribución de la variable es ***simétrica***. En caso contrario, hay que evaluar en qué dirección se presenta la asimetría. Además, es importante evaluar cuál es la magnitud de las cajas y de las colas, pues cuanto más cercanos sean los valores, habrá menor variabilidad en ella.

Se basa en medidas de posición; más específicamente, utiliza la mediana, el percentil 25 y el percentil 75 para su creación.

A partir de estos últimos, calcula ***el rango intercuartílico***, el cual es la medida de dispersión de la variable.

Para la construcción del gráfico de caja y bigote, se ubican inicialmente los valores correspondientes a los percentiles 25 y 75, los cuales le dan forma a la caja. La mediana va a dividir esta caja en dos partes y va a ser el valor de referencia del resto de la gráfica.

Posteriormente, se calcula el rango intercuartílico, que se utiliza para construir los bigotes; en el caso del bigote inferior, se toma el valor correspondiente al percentil 25 y se le resta 1,5 veces el rango intercuartílico. Este resultado se compara con los valores mínimos observados en la variable, y se identifica el valor que sea más cercano, pero no menor, que el valor antes calculado; los valores que sean inferiores a dicho valor corresponden a datos atípicos, los cuales se presentan, por lo general, con un círculo o un asterisco.

En el caso del bigote superior, se calcula tomando el valor del percentil 75, pero, a diferencia del anterior, se le suma 1,5 veces el rango intercuartílico; nuevamente, se identifica el valor observado más cercano que no sea mayor que el valor calculado; los valores que queden por encima de ese valor también corresponden a datos atípicos. La **figura 33.6** presenta el gráfico de caja y bigotes de la variable pulsaciones por minuto.

Puede verse en la **figura 33.6** que la distribución no es simétrica, pues la distancia que hay entre los percentiles y la mediana es muy diferente, y, además, hay mucha variabilidad; especialmente, en los valores mayores que la

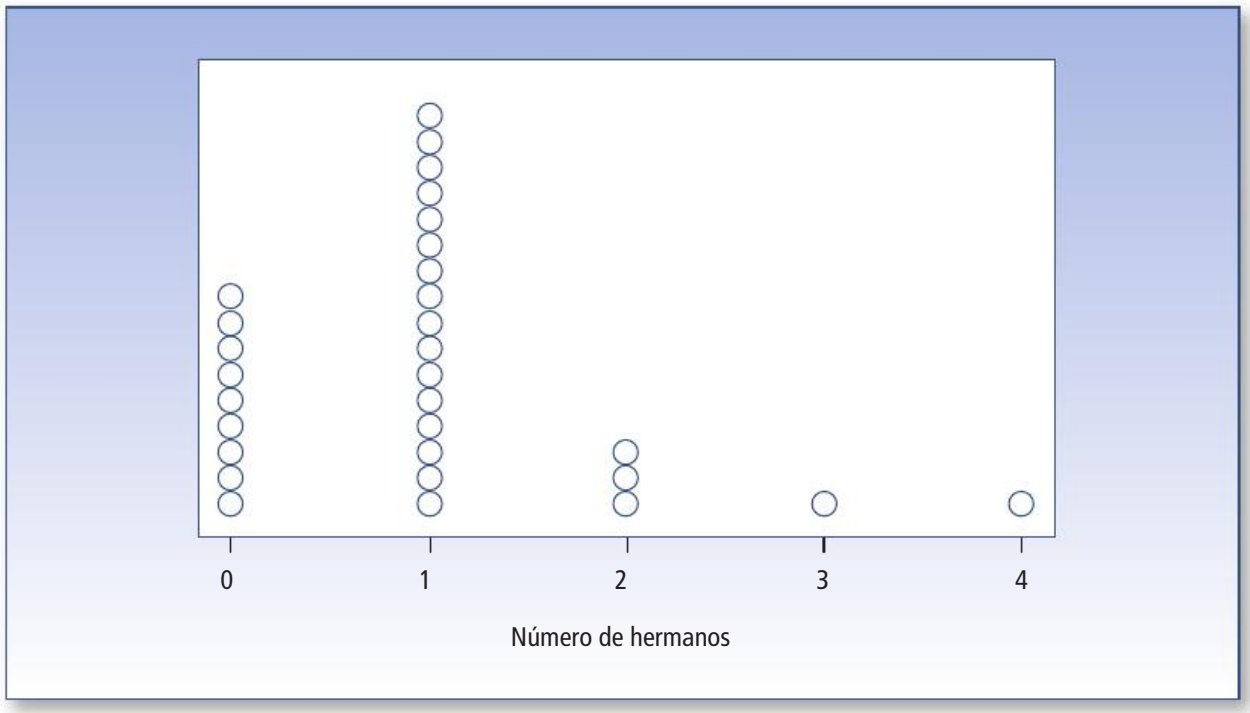


Figura 33.5 **Gráfico de punto para mostrar la información "Distribución del número de pulsaciones de 30 recién nacidos en un** hospital en 2014".

mediana. Como información adicional, debe recordarse que los valores mínimo y máximo de esta variable correspondían a 71 y 189, respectivamente. Si se evalúan los valores de los bigotes, puede verse que coinciden con dichos valores; por lo tanto, en este ejemplo no se detectaron valores atípicos, y, por ende, no se presentan círculos ni asteriscos para identificarlos.

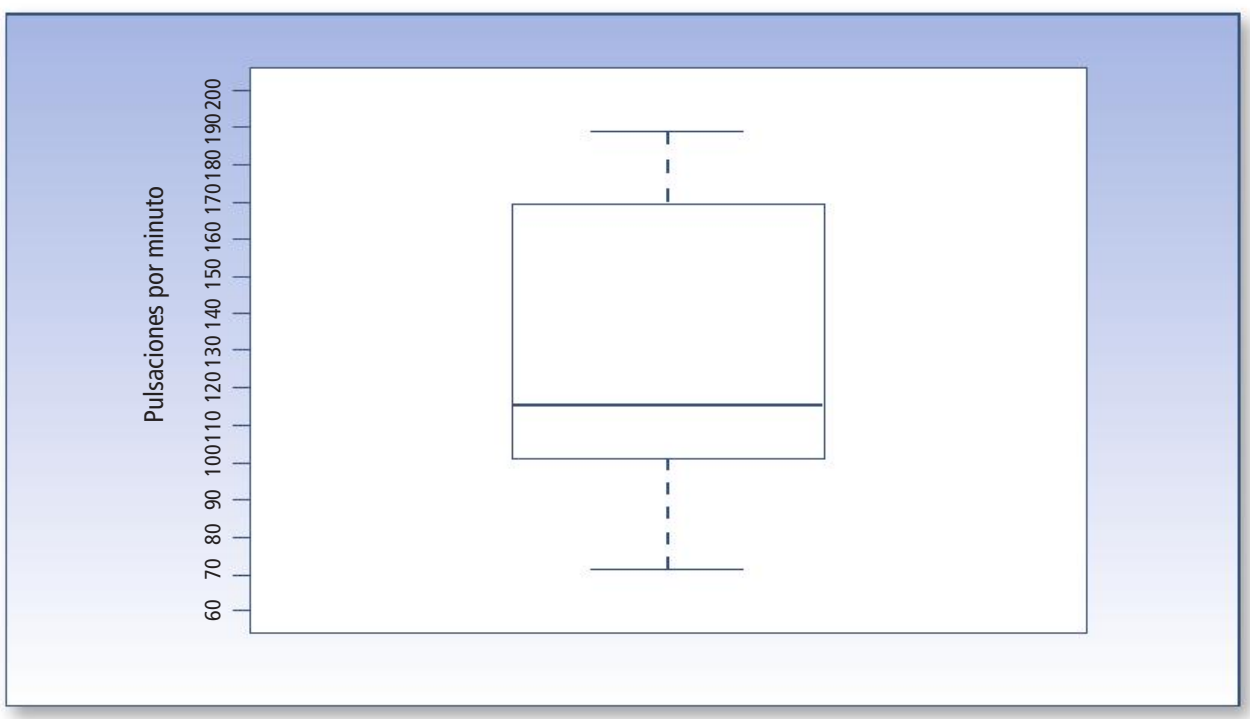


Figura 33.6 **Gráfico** *Box plot* para mostrar la información "Número de pulsaciones de 30 recién nacidos en un hospital en 2014".

Gráficos de caja de letras y valores

Los gráficos de caja de letra y valores se pueden considerar una extensión de los gráficos de caja y bigotes, pero en vez de tener una caja dividida en dos partes iguales, este gráfico busca, de manera recursiva, construir más de dos cajas para mostrar de forma más detallada la distribución de los datos[16].

Aunque sirven para evaluar la simetría y la dispersión de una variable, en general se recomiendan para tamaños de muestra muy grandes, pues se considera que en esta situación, los gráficos de caja y bigotes pueden ser insuficientes, y, en algunos casos, poco informativos.

En especial, al momento de identificar datos atípicos. Para esto, el gráfico clasifica los valores atípicos en dos categorías: ***datos atípicos extremos*** y *datos atípicos remotos*.

La letra se utiliza para ir definiendo la medida de posición que se va calculando en cada partición, de manera ordenada. Por ejemplo, en el caso de la caja con bigotes, que es un caso particular de estos, la primera medida de posición que se calcula es la ***mediana***, la cual divide el conjunto de datos en dos grupos de igual magnitud; generalmente, a esta partición se le asigna la letra M (de mediana).

La segunda partición corresponde a dividir cada uno de los grupos creados por la mediana en dos partes iguales; o sea, divide el total de datos en cuatro partes iguales (también conocidos como ***cuartiles***), por lo cual se les asignará la letra F, que corresponde a la inicial de cuatro en inglés (***four***).

La siguiente partición corresponde a la partición de los nuevos grupos, que resulta de dividir cada uno de ellos en dos partes iguales; o sea, al final tenemos 8 nuevos grupos (***octiles***), y se le asignaría la letra E, que corresponde a la letra inicial de ocho en inglés (***eight***). Este proceso se repite hasta obtener el número de grupos que sea necesario según el tamaños de la muestra (es decir, 16, 32, 64, etc.). Para estos casos, por facilidad, se van asignando letras en orden alfabético descendente, y tomando como inicio la letra antes establecida (E); es decir, para 16 corresponde la letra D; para 32, la letra C, y así sucesivamente. En la **tabla 33.6** se presenta un bosquejo de cómo se hace la asignación de las letras.

Una vez creadas las cajas, se procede a calcular las ***fronteras***: la frontera interior y la frontera exterior. La frontera interior corresponde a los bigotes calculados en el gráfico de caja y bigote. Las fronteras exteriores se deben calcular tanto para los valores inferiores como para los valores superiores, de la siguiente manera:

$$\textit{Frontera Exterior Superior: } F_{sup} + 3*(F_{sup} - F_{inf})$$
$$\textit{Frontera Exterior Inferior: } F_{inf} + 3*(F_{sup} - F_{inf})$$

Este gráfico no presenta los valores atípicos que quedan entre la frontera interior y la frontera exterior (estos se pueden apreciar en un gráfico de ***box plot***), pero sí, a los valores que quedan por fuera de la frontera exterior, los cuales son llamados ***valores atípicos remotos***. En la **figura 33.7** se presenta un gráfico de caja de letras y valores para la variable en análisis.

Tabla de frecuencias para una variable cuantitativa continua

En el caso de las variables continuas, como la variable presenta muchas opciones de respuesta, también se debe especificar una categoría o un intervalo de clase; por lo tanto, también se deben conocer el número y la amplitud de estas categorías. Como en el caso de las variables discretas para valores finitos, el investigador puede elegir entre dos posibilidades: 1) determi-

Tabla 33.6 Bosquejo de asignación de las letras en el gráfico de caja de letras y valores

Número de grupos	Letra
2	M
4	F
8	E
16	D
32	C
64	B
128	A
256	Z
512	Y
…	…

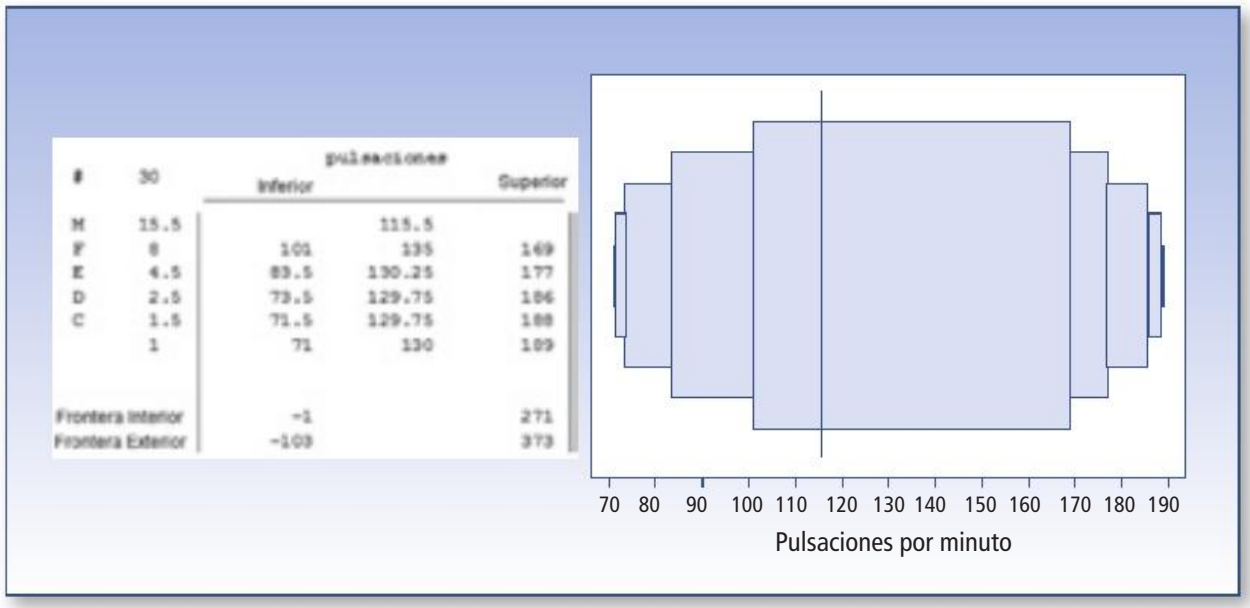


Figura 33.7 Gráfico de caja de letras y valores para mostrar la información "Número de pulsaciones de 30 recién nacidos en un hospital en 2014".

nar a su propia conveniencia cuántas categorías construir, o 2) utilizar las estrategias o las reglas ya mencionadas.

Si se toma como ejemplo la variable peso al nacer, al aplicar las reglas de Sturges y de Lane, estas no cambian respecto a las calculadas para la variable pulsaciones, pues el tamaño de la muestra es el mismo. La diferencia va a estar en las otras reglas, ya que las medidas de variabilidad y dispersión sí cambian. Al calcular el número de categorías para esta nueva variable, se encuentra que tanto para Scott como para Freedman-Diaconis se obtienen tres categorías, mientras para Doane se obtienen seis.

Al calcular la amplitud del intervalo, este corresponde al rango ($R = 3900 - 1060 = 2840$) dividido por el número de categorías ($k = 3$), y se obtiene una amplitud de 947. Por tratarse de una variable continua, se tiene que donde termina una categoría comienza la otra; por lo tanto, para efectos de una mejor visualización y una mejor comprensión de la tabla, se han incluido los valores de los límites de las categorías en paréntesis (cuadrado para el límite inferior y redondo para el límite superior).

Esta notación se usa para advertir si un determinado valor entra o no en una de esas categorías, de la siguiente manera: el paréntesis cuadrado indica que el valor dado se incluye en la *categoría*, mientras el redondo se toma como *referencia*, pero el intervalo nunca lo va a incluir, por ejemplo: [1060-2007], e indica que ese intervalo incluye dentro de sus valores a 1060, pero nunca el valor de 2007, pues su límite va a llegar hasta el valor 2006,999. En la **tabla 33.7** se presentan las respectivas categorías y su correspondiente distribución de frecuencias.

Muchas veces, luego de construida la tabla se ve que los intervalos de clase establecidos no siguen una lógica clínica, o bien, son difíciles de interpretar. En tales escenarios, se puede recurrir al uso de categorías a conveniencia del investigador, con el número de categorías recomendado.

Tabla 33.7 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del peso al nacer de los 30 recién nacidos en un hospital en 2014

Peso al nacer	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
[1060-2007)	**8**	**26,7**	**8**	**26,7**
[2007-2954)	11	36,67	19	63,3
[2954-3901)	11	36,67	30	100,0

Por ejemplo, para la variable peso al nacer, tiene más sentido clínico hacer la tabla con categorías de peso utilizadas internacionalmente, como se presenta en la **tabla 33.8.**

Gráficos para una variable cuantitativa continua

Hay una gran variedad de gráficos para presentar visualmente datos numéricos de tipo continuo. Dentro de los gráficos más utilizados están: histogramas, polígonos de frecuencias, gráficos de tallo y hojas, gráficos de caja y bigote (***box plot***) y gráficos de caja de letras y valores[13-15,17,18].

Histograma de frecuencias

El histograma es un gráfico similar al de barras simples, que busca presentar información calculada a través de las tablas de frecuencias para variables continuas.

Su diferencia principal radica en que, dada la continuidad de sus valores, las barras no están separadas, se encuentran ubicadas manteniendo el orden de sus categorías y su tamaño es proporcional a la frecuencia de los valores presentados. Este último aspecto es muy importante, pues uno de los errores más comunes es asumir igual tamaño para sus barras, situación que es equivocada cuando la amplitud de los intervalos es diferente; en tal situación, se debe modificar la altura de la barra, de manera que su nueva área represente la frecuencia real de la categoría. La variable de interés siempre va a estar representada en el eje horizontal, y la frecuencia o el porcentaje, en el eje vertical.

Se utiliza, principalmente, para dar una idea visual de la forma de la distribución de la variable; por lo tanto, es útil para detectar si la distribución es simétrica o no, y muy a menudo se utiliza para evaluar la normalidad de esta. La **figura 33.8 presenta un histograma de frecuencias** para la variable peso al nacer, considerando: a) categorías de igual amplitud y b) categorías de diferente amplitud, agrupando la primera y la segunda categoría.

Polígono de frecuencias

El caso del polígono de frecuencias es una modificación del histograma de frecuencias, y busca ver de una manera más clara la forma de la distribución de la variable. Para ello, en vez de dibujar barras, lo que hace es identificar cuál es el punto medio de cada uno de los intervalos (***marca de clase***), **y los une a través de una línea** recta. Una de las características de este gráfico es que debe comenzar y terminar con una frecuencia de 0; para ello, construye categorías de igual amplitud: una antes de la primera categoría y otra después de la última categoría, y, nuevamente, en cada una de ellas identifica la marca de clase, y las une, finalmente, con las anteriores.

Una de sus principales desventajas es que es más difícil de interpretar cuando los tamaños de los intervalos de clase son de diferente amplitud. En la **figura 33.9 se presenta un polígono de frecuencias** para la variable peso al nacer.

Gráfico de tallo y hojas

Al igual que en el caso de histogramas, la idea de un gráfico de tallo y hojas es ordenar y organizar los datos en grupos, de tal manera que se pueda identificar la forma de distribución de la variable, pero en términos más sencillos. Para eso, crea grupos de valores (llamados ***tallos***), **que, en general, se van a componer, según** la variable, en análisis de unidades, decenas, centenas, etc.; y una vez creados, evalúa qué valores específicos caen dentro de cada una de

Tabla 33.8 Distribución de frecuencias absolutas y porcentuales (simples y acumuladas) del peso al nacer de los 30 recién nacidos en un hospital en 2014

Peso al nacer	Frecuencias absolutas	Porcentaje	Frecuencia absoluta acumulada	Porcentaje acumulado
[1000-2000)	8	26,7	8	26,7
[2000-3000)	12	40,0	20	66,7
[3000-4000)	10	33,3	30	100,0

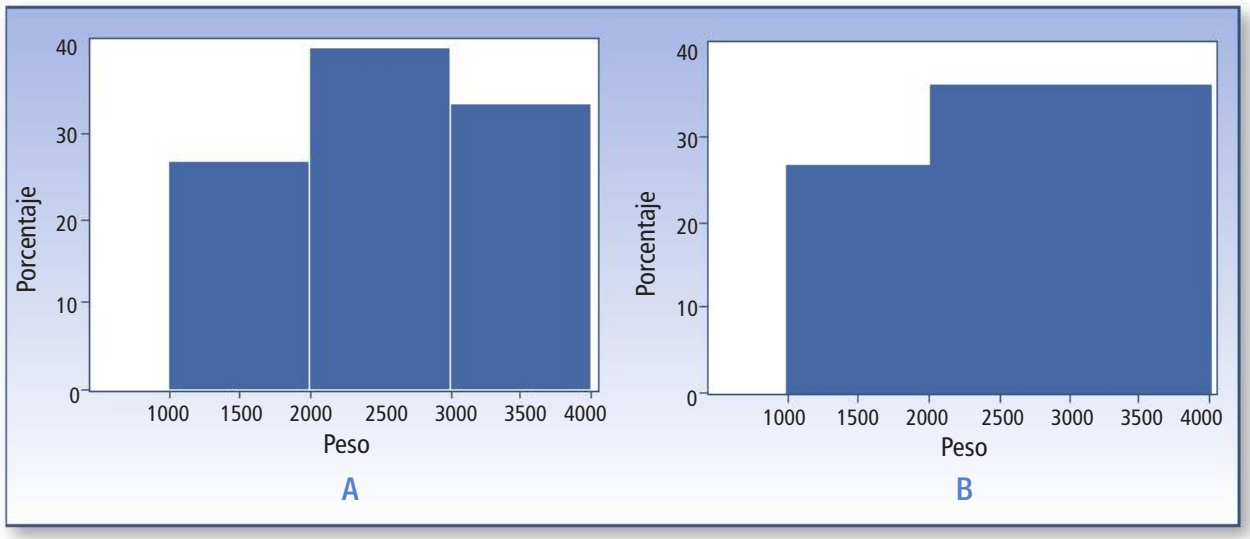


Figura 33.8 Histogramas de frecuencias para representar la información "Peso al nacer de los 30 recién nacidos en un hospital en 2014".

ellas (llamadas ***hojas***). Para una mejor comprensión, se va a ir explicando su construcción tomando como ejemplo la variable estancia hospitalaria.

El primer paso, lo anterior es ordenar los datos de menor a mayor, e identificar el rango en el que estos varían, con el fin de decidir cuál va a ser el criterio para crear los grupos. Los siguientes son los valores: 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 4, 4, 4, 5, 6, 6, 6, 8, 8, 9, 11, 11, 16, 19, 19, 27, 47, 54, 79. Como puede verse, son valores que oscilan entre 1 y 79 días; por lo tanto, pueden tomarse como tallo las decenas, y así crear los siguientes grupos para el tallo: 0, 1, 2, 3, 4, 5, 6 y 7. Se ubican estos valores en la primera columna, uno debajo del otro; y para poner las hojas, se van colocando en orden ascendente los dígitos correspondientes a cada una de las

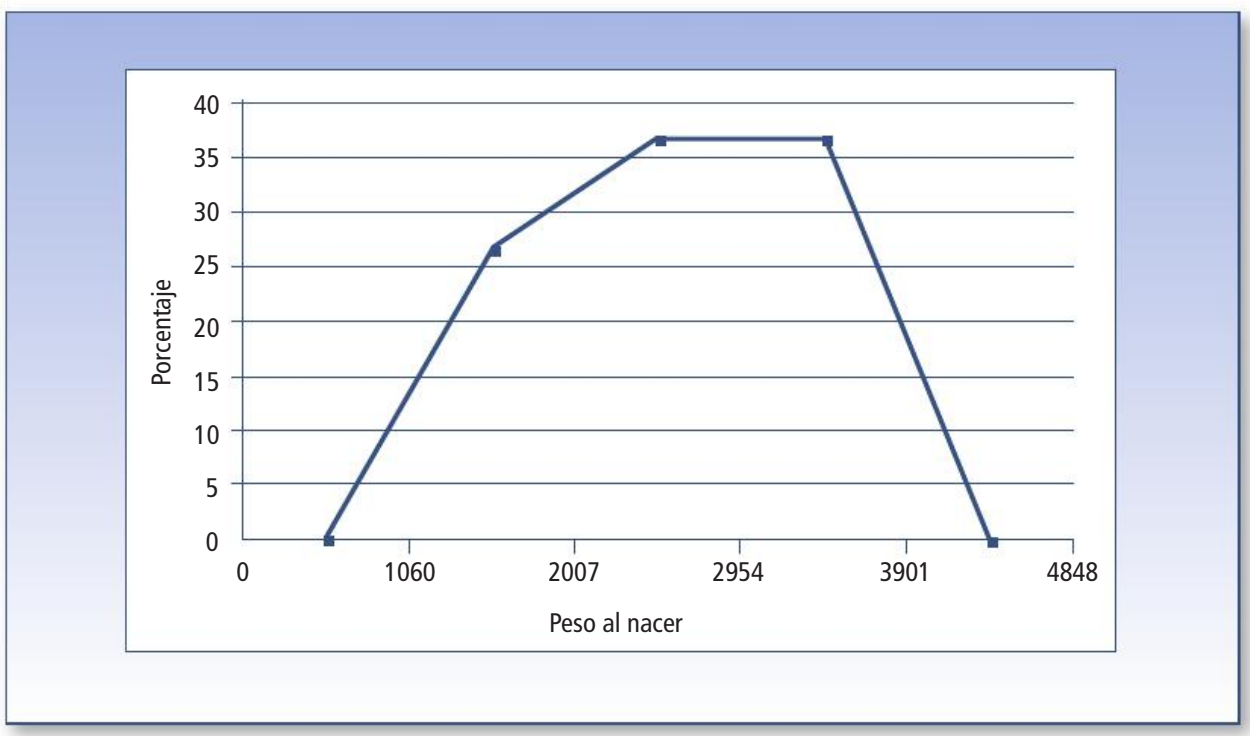


Figura 33.9 Histogramas de frecuencias para representar la información "Peso al nacer de los 30 recién nacidos en un hospital en 2014".

decenas. Así, por ejemplo, el valor 1 representa valores entre 10 y 19; por lo tanto, al frente del 1 se colocarán en orden los valores 1, 1, 6, 9 y 9. El gráfico de tallo y hojas final se presenta en la **figura 33.10**.

Gráfico de línea

El gráfico de línea es otra manera de visualizar variables continuas; como en el caso de los histogramas, se recomienda cuando se quieren mostrar diferencias en las frecuencias o los porcentajes entre distintas categorías de una variable continua, pero, especialmente, para identificar distribuciones y tendencias. La frecuencia va a estar representada por un punto, el cual se calcula en cada una de las categorías, y los cuales, finalmente, son unidos a través de una línea recta. El polígono de frecuencias puede considerarse un caso particular de este tipo de gráficos.

En general, los gráficos de línea son más apropiados para evaluar la tendencia o el comportamiento de una variable; en especial, cuando están en función del tiempo. En la **figura 33.11** se presenta un ejemplo con la distribución de frecuencias de la estancia hospitalaria.

Tablas y gráficos para dos variables

En el caso de analizar dos variables, el enfoque del análisis se centra, principalmente, en mirar la relación que hay entre ellas.

Un primer escenario es cuando las dos variables son categóricas. En tal caso, las relaciones se estudian mediante tablas de contingencia (también conocidas como ***tablas cruzadas*** o *tablas de doble entrada*) o a través de gráficos específicos utilizados para tal fin.

0	1 1 1 2 2 2 2 2 2 2 3 4 4 4 5 6 6 6 8 8 9
1	1 1 6 9 9
2	7
3	
4	7
5	4
6	
7	9

Figura 33.10 Gráfico de tallo y hojas para que representa la información "Estancia hospitalaria de 30 recién nacidos en un hospital en 2014".

Existe también la posibilidad de tener una variable cualitativa y una cuantitativa o dos variables cuantitativas. En el primer caso, se puede tener como estrategia crear tablas de frecuencias de la variable cuantitativa (***categorizar la variable***) y emplear para su análisis tablas de contingencia o sus gráficos correspondientes; también se puede hacer una exploración más visual a través de gráficos tales como el ***box plot*** y la pirámide poblacional.

En el caso de dos variables cuantitativas, no tiene mucho sentido hacer tablas de frecuencias de los valores originales, por lo que, generalmente, se recurre a utilizar alguna de las dos estrategias anteriores; o sea, categorizar una o las dos variables. En este escenario, tiene más sentido hacer un análisis a través de gráficos tales como los gráficos de dispersión y los gráficos de línea.

Tablas para dos variables cualitativas

Al igual que para el caso de una variable, en las tablas de contingencia de variables nominales y ordinales la información que se presenta corresponde a la distribución de las frecuencias absolutas relativas (o ***porcentajes***). A diferencia de las tablas para una variable, las tablas de contingencia contienen dentro de las ***filas*** las categorías de respuesta de una de las variables, y dentro de las ***columnas***, las categorías de respuesta de la otra variable.

No es raro hacer una distinción entre las variables; especialmente, cuando se quiere evaluar si una variable independiente (***de exposición***) tiene asociación a un desenlace de interés (o *respuesta*). En tales casos, se recomienda establecer las categorías de la variable respuesta en la parte de las columnas, y las de la exposición, en las filas. En caso de no haber tal distinción, se recomienda poner en las filas la variable que presente el mayor número de opciones de respuesta.

La información que más se presenta corresponde a las frecuencias absolutas, lo cual, simplemente, consiste en contar el número de casos que sucede para cada combinación de las categorías de las variables en análisis. De manera opcional, se puede incluir en dichas tablas una

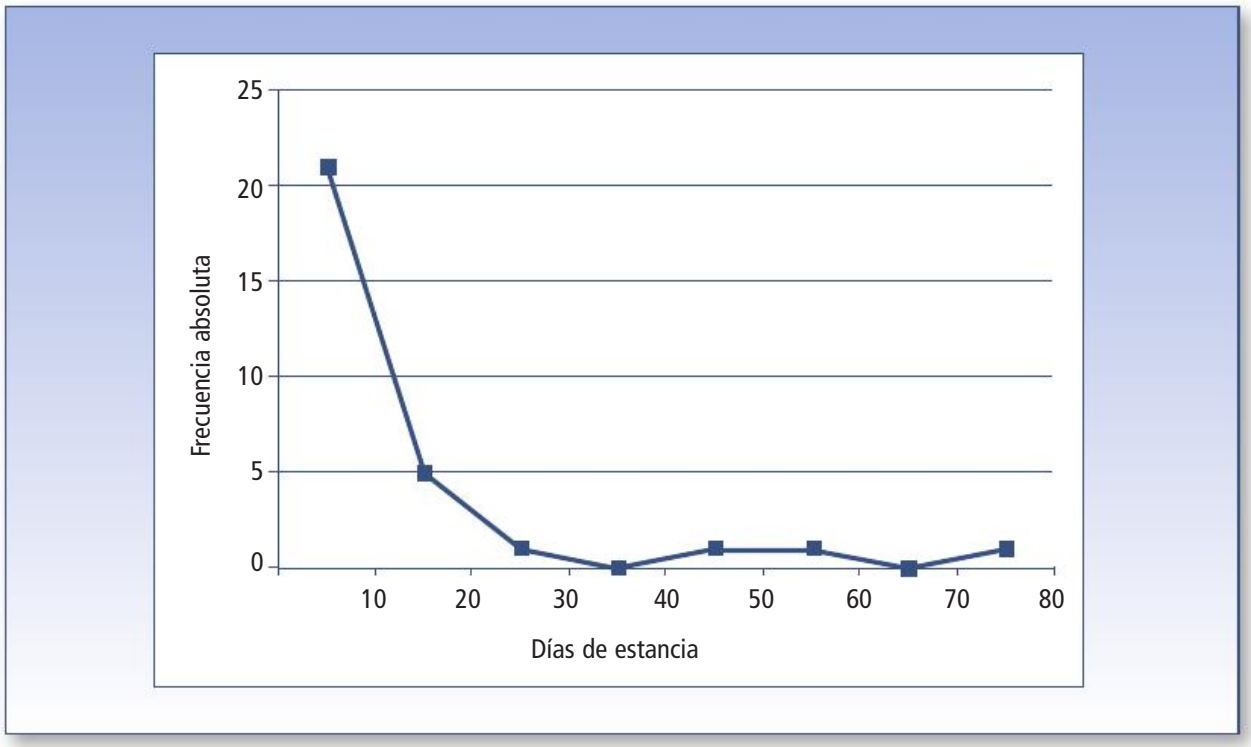


Figura 33.11 Gráfico de línea para representar la información "Estancia hospitalaria de 30 recién nacidos en un hospital en 2014".

columna y una fila adicional, donde se va a incluir la información correspondiente a las frecuencias absolutas referentes al análisis univariado de cada una de ellas.

La diferencia principal con una tabla para una variable única radica en el cálculo de las frecuencias relativas o de los porcentajes, pues al tener una tabla de doble entrada se pueden condicionar los resultados en relación con las filas, con las columnas o con el total de observaciones. No se recomienda incluir los tres en la misma tabla, ya que pueden llegar a generar confusión a los lectores, por lo cual se recomienda tener claro cómo se quieren presentar los resultados, y únicamente colocar las proporciones o los porcentajes que sean de interés, de tal forma que sea más clara su interpretación.

Por ejemplo, para evaluar si hay asociación entre el estado de salud al nacer de los neonatos y el sexo, se va a construir una tabla de frecuencias. Se ubican en sus filas las categorías del estado de salud, y en sus columnas, las relacionadas con el sexo.

En las **tablas 33.9** y **33.10** se presentan los resultados de las frecuencias absolutas encontradas, para evaluar esta asociación considerando el porcentaje fila y el porcentaje columna, respectivamente. El valor que aparece en la celda correspondiente a estado de salud bueno y sexo masculino de la **tabla 33.9** se interpreta respecto a esa fila; es decir, de los recién nacidos que presentaron un estado de salud bueno, el 40,9 % eran de sexo masculino. En cuanto al valor que aparece en la celda correspondiente a estado de salud bueno y sexo masculino de la tabla 10, se interpreta respecto a esa columna; es decir, de los recién nacidos de sexo masculino, el 81,8 % presentaron un estado de salud bueno.

Gráficos para dos variables cualitativas

A partir de la tabla de frecuencias antes construida, se pueden considerar algunos gráficos específicos para su presentación. Dentro de los gráficos más utilizados para presentar este tipo de datos se tienen gráficos de barras agrupadas y los gráficos de barras apiladas.

Gráficos de barras agrupadas

Este tipo de gráfico se utiliza para comparar entre sí distintos grupos o variables cualitativas. Se basa en las tablas de asociación antes calculadas, por lo tanto, tiene las mismas recomendaciones. A diferencia del gráfico de barras simples, dibuja para cada categoría de una variable el total de barras presentadas en

Tabla 33.9 Tabla de frecuencias absolutas y porcentajes por fila del estado de salud al nacer y el sexo de 30 recién nacidos en un hospital en 2014

Estado de salud al nacer	Sexo		Total n (%)
	Masculino *n* (%)	Femenino *n* (%)	
Bueno	9 (40,9)	13 (59,1)	22 (100,0)
Regular	**0 (0,0)**	**5 (100,0)**	**5 (100,0)**
Malo	2 (66,7)	1 (33,3)	3 (100,0)
Total	11 (36,7)	19 (63,3)	30 (100,0)

Tabla 33.10 Tabla de frecuencias absolutas y porcentajes por columna del estado de salud al nacer y el sexo de 30 recién nacidos en un hospital en 201

Estado de salud al nacer	Sexo		Total *n* (%)
	Masculino *n* (%)	Femenino *n* (%)	
Bueno	9 (81,8)	13 (68,4)	22 (73,3)
Regular	0 (0,0)	5 (26,3)	5 (16,7)
Malo	2 (18,2)	1 (5,3)	3 (10,0)
Total	11 (100,0)	19 (100,0)	30 (100,0)

la segunda categoría; es decir, para el ejemplo anterior, si se le mira desde las ***filas***, podrían graficarse tres barras (correspondientes a las categorías bueno, regular y malo) para el sexo masculino, y tres, para el sexo masculino; o si se lo mira desde las ***columnas***, podrían graficarse dos barras (que corresponden a sexo masculino y sexo femenino) para estado de salud bueno, dos barras para estado de salud regular y otras dos barras para estado de salud malo **(figura 33.12)**.

Gráficos de barras apiladas

Este tipo de gráfico es equivalente al gráfico de barras agrupadas. La diferencia entre ambos estriba en que para cada categoría se ubica la información de las categorías de la segunda variable en una misma barra. Por ejemplo, si se mira desde las columnas, puede apilarse en una misma barra la información correspondiente al sexo de los recién nacidos (masculino y femenino) para cada estado de salud **(figura 33.13)**.

Gráficos para una variable cualitativa y una cuantitativa

La idea de estos gráficos es evaluar la distribución de una variable continua de interés según los grupos de una segunda variable. Su propósito principal es comparar entre sí los grupos de la variable categórica. Los polígonos de frecuencias están indicados en esta situación, así como las pirámides poblacionales y los gráficos de caja y bigotes.

Pirámides poblacionales

Las pirámides poblacionales son una forma especial de un histograma, y en las que se busca representar de manera conjunta la distribución de los grupos de edad de una población, según el sexo.

Inicialmente, se hace un histograma de frecuencias para los hombres, y otro para las mujeres, y luego se ubican de tal manera que sus barras queden en posición horizontal, y contrapuestas la una con la otra. En la gráfica de la izquierda, el valor inicial (generalmente, 0) está a la derecha, en tanto que para la de la derecha el valor 0 está a la izquierda.

La forma de pirámide era evidente cuando la mayoría de la población era menor de edad, y seguía una tendencia decreciente a medida que aumentaba la edad. Sin embargo, en la ac-

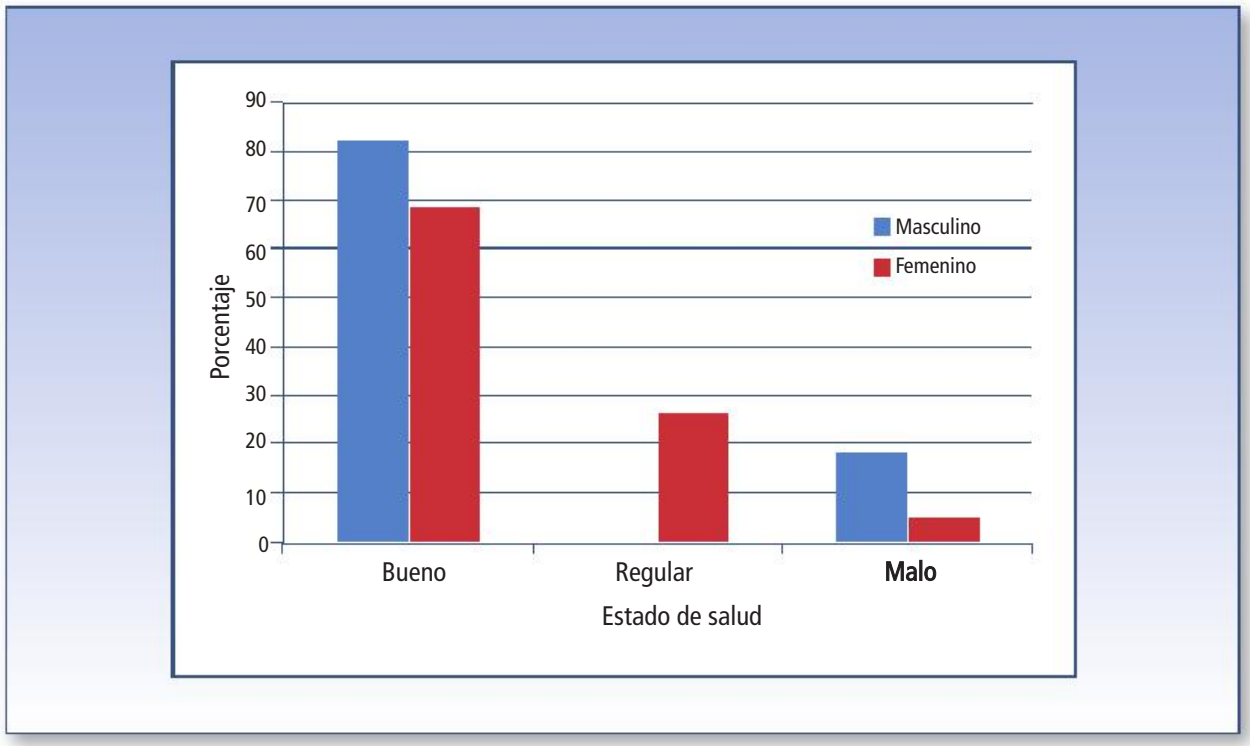


Figura 33.12 Gráfico de barras agrupadas para mostrar la información "Estado de salud y sexo de 30 recién nacidos en un hospital en 2014".

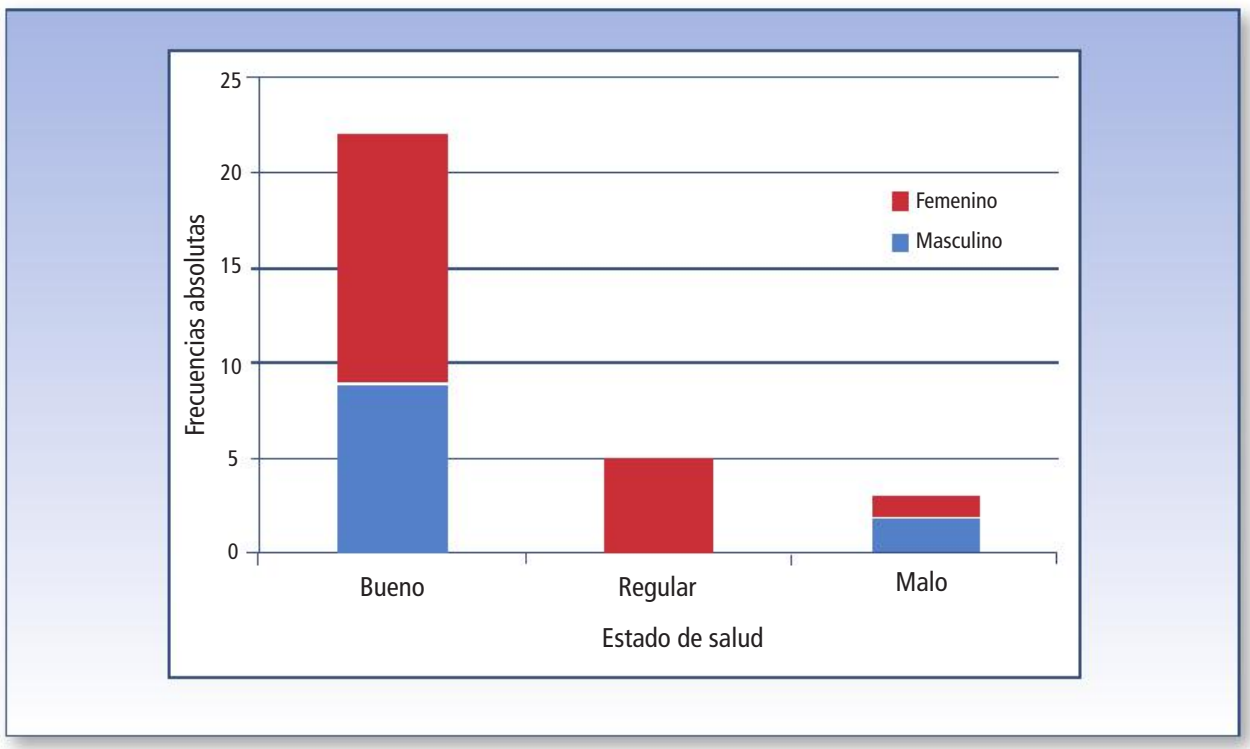


Figura 33.13 Gráfico de barras apiladas para mostrar la información "Estado de salud y sexo de 30 recién nacidos en un hospital en 2014".

tualidad dicha tendencia ha estado cambiando por el aumento de la supervivencia en mayores y por la disminución de la natalidad, y tiene más interés evaluar las variaciones en el comportamiento de la composición poblacional a través de esta pirámide, la cual, incluso, puede, en algunos casos, ser una pirámide invertida **(figura 33.14)**.

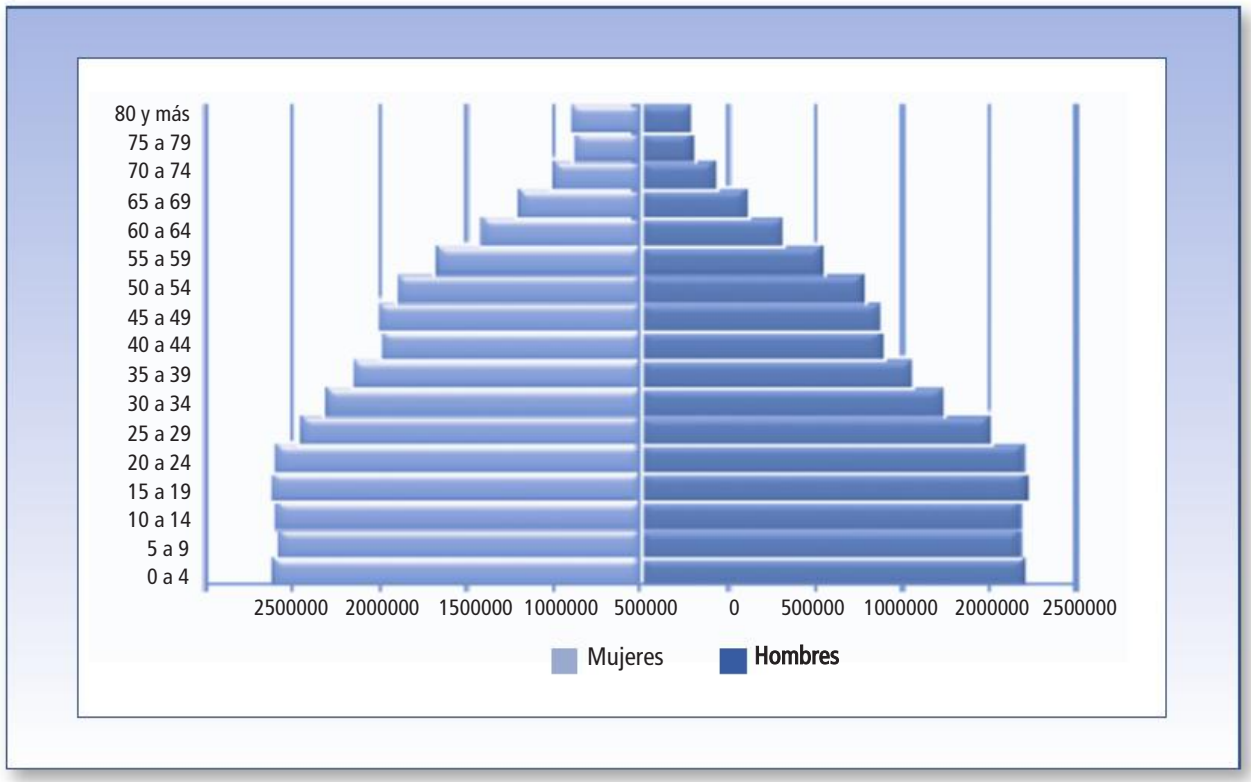


Figura 33.14 Pirámide poblacional para mostrar la información "Población colombiana proyectada para el año 2015".
Fuente: Departamento Administrativo Nacional de Estadística (DANE)[19].

En el caso de las cajas con bigotes **(figura 33.15)** y de los polígonos de frecuencia, la estrategia que debe seguirse es la misma que para el caso de una variable, pero se hace independientemente para cada categoría de la variable categórica.

Gráficos para dos variables cuantitativas

Entre los gráficos más utilizados están los ***gráficos de dispersión*** y los ***gráficos de línea***[20].

En el caso de los gráficos de dispersión, estos buscan evaluar si hay asociación entre dos variables cuantitativas.

Para ello, se dibujan puntos que representan un par de valores de dos variables evaluadas. Por ejemplo, si se quiere evaluar la relación entre el peso al nacer y la edad gestacional, un punto representa la información de ambas medidas en un recién nacido en particular. La **figura 33.16** presenta el gráfico de dispersión para estas dos variables.

Gráficos para más de dos variables cuantitativas

Este tipo de gráfico es recomendable cuando se tienen más de dos variables de interés que se quieren evaluar simultáneamente. Este tipo de análisis se usa, especialmente, en el análisis multivariado de datos. Los gráficos más utilizados son las caras de Chernoff, los gráficos de red y los gráficos de estrella[13,15].

Caras de Chernoff

Este otro tipo de gráfico multivariado se caracteriza por representar la información hasta de 18 variables a la vez, en forma de un rostro humano.

Cada una de las variables va a estar representada por una de las partes que componen el rostro; es decir, el tamaño, la forma, la ubicación y la orientación de los ojos, los oídos, la boca, la nariz y las cejas, así como la forma de la cara, van a estar representando la magnitud de los valores de cada una de las variables.

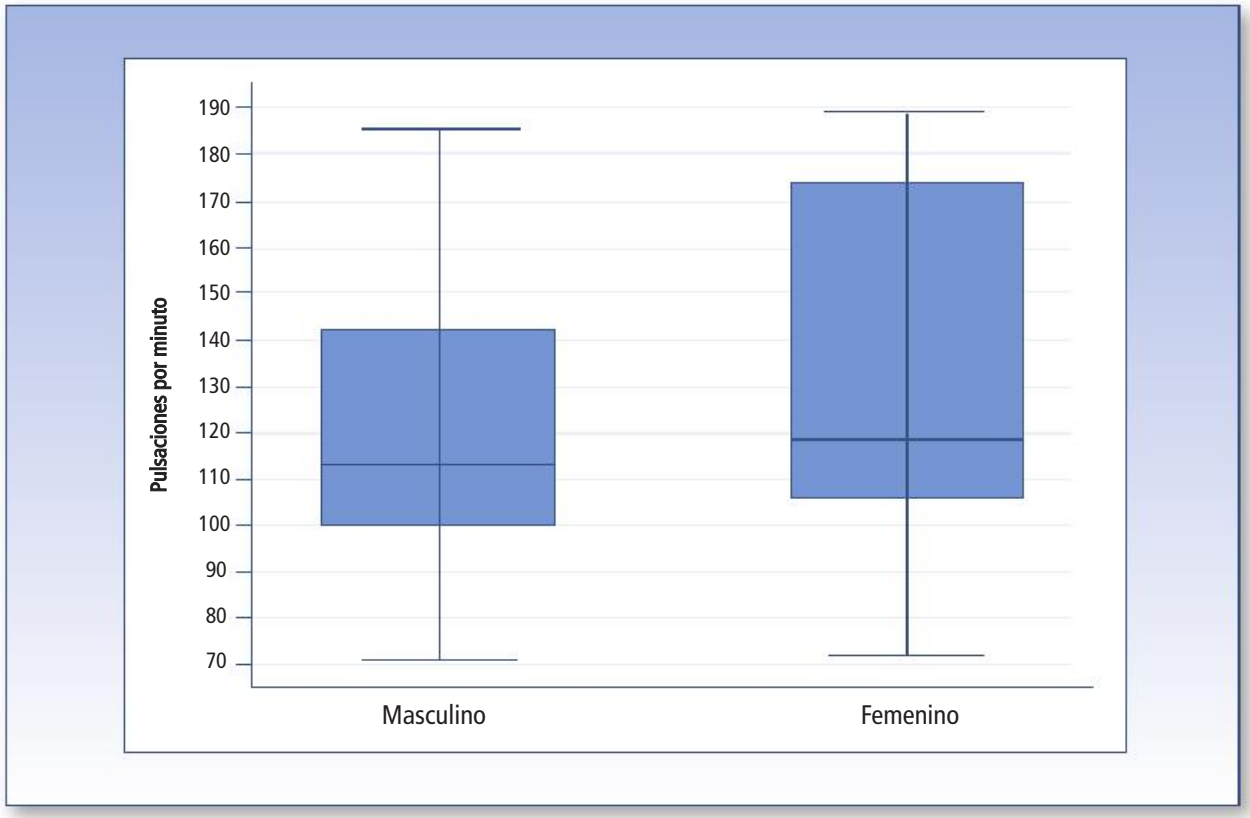


Figura 33.15 Gráfico de caja y bigotes para mostrar la información "Pulsaciones por minuto según el sexo de 30 recién nacidos en un hospital en 2014".

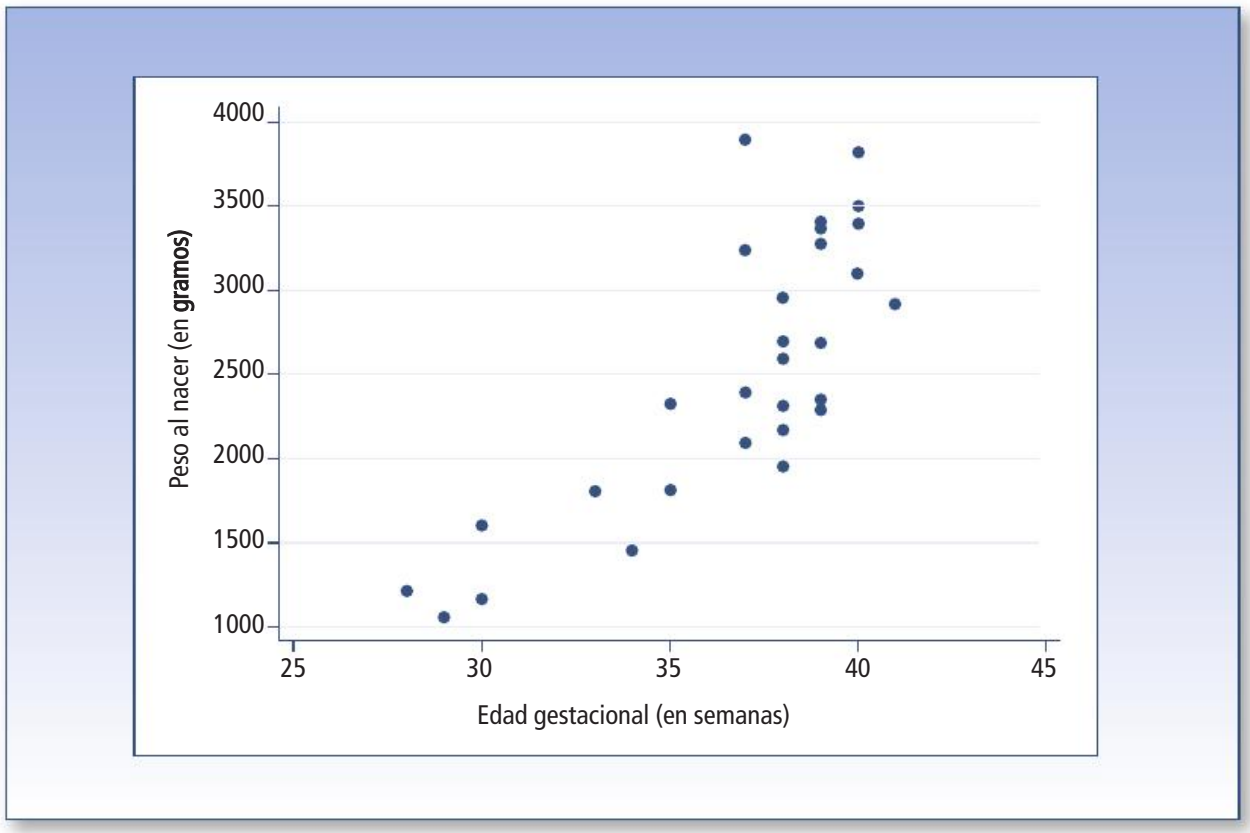


Figura 33.16 Gráfico de dispersión para mostrar la comparación entre "Peso al nacer y edad gestacional de 30 recién nacidos en un hospital en 2014".

Al tratarse de rostros humanos, se espera que se puedan reconocer fácilmente pequeños cambios o patrones, y permite, incluso, llegar a ver asociaciones entre algunas de las variables.

Su principal desventaja es que no es recomendable para representar información de tamaños de muestra grandes, pues cada rostro va a estar representando a un individuo. Para efectos de una mejor comprensión, la **figura 33.17** presenta la información de los primeros 9 recién nacidos de nuestro ejemplo base.

Gráficos de red

Un gráfico de red, también conocido como *gráfico de telaraña* o *gráfico de radar* **(figura 33.18), es un método para representar datos de** tres o más variables numéricas a través de ejes o radios que tienen un punto de partida común para todas. La idea es que se va a crear un radio por cada individuo en análisis, por lo cual resulta adecuado cuando el tamaño de la muestra es pequeño o moderado. Su principal característica es tener todos los radios equidistantes; o sea, el ángulo entre ellos debe ser el mismo.

La longitud de los radios va a depender del máximo valor observado entre todas las variables; por lo tanto, es muy recomendado para variables ordinales (ojalá, con las mismas opciones de respuesta). En el caso de variables continuas con magnitudes muy diferentes (como, por ejemplo, peso al nacer y edad gestacional), los valores de la segunda van a ser difíciles de identificar, ya que la magnitud de los ejes va a ser muy grande, al depender del peso al nacer.

A diferencia de las caras de Chernoff, cada forma geométrica dentro de la telaraña va a representar la información de una variable, y su forma depende de los valores observados para esa variable por cada uno de los individuos, y los conecta a través de una línea recta entre cada uno de los radios.

Son adecuados principalmente, para identificar valores atípicos y homogeneidad, y son muy recomendados especialmente, para variables ordinales; ojalá, con las mismas opciones de respuesta.

Gráficos de estrella

Son otro de los métodos con forma geométrica que sirven para representar datos de tres o más variables numéricas a través de ejes o radios, y son parecidas al gráfico de radar (por lo que ambos tienden a ser confundidos). Sin embargo, a diferencia del anterior, los gráficos de estrella definen el número de ejes por el número de variables en análisis; es decir, cada estrella representa a un individuo y los ejes de la estrella representan cada una de las variables,

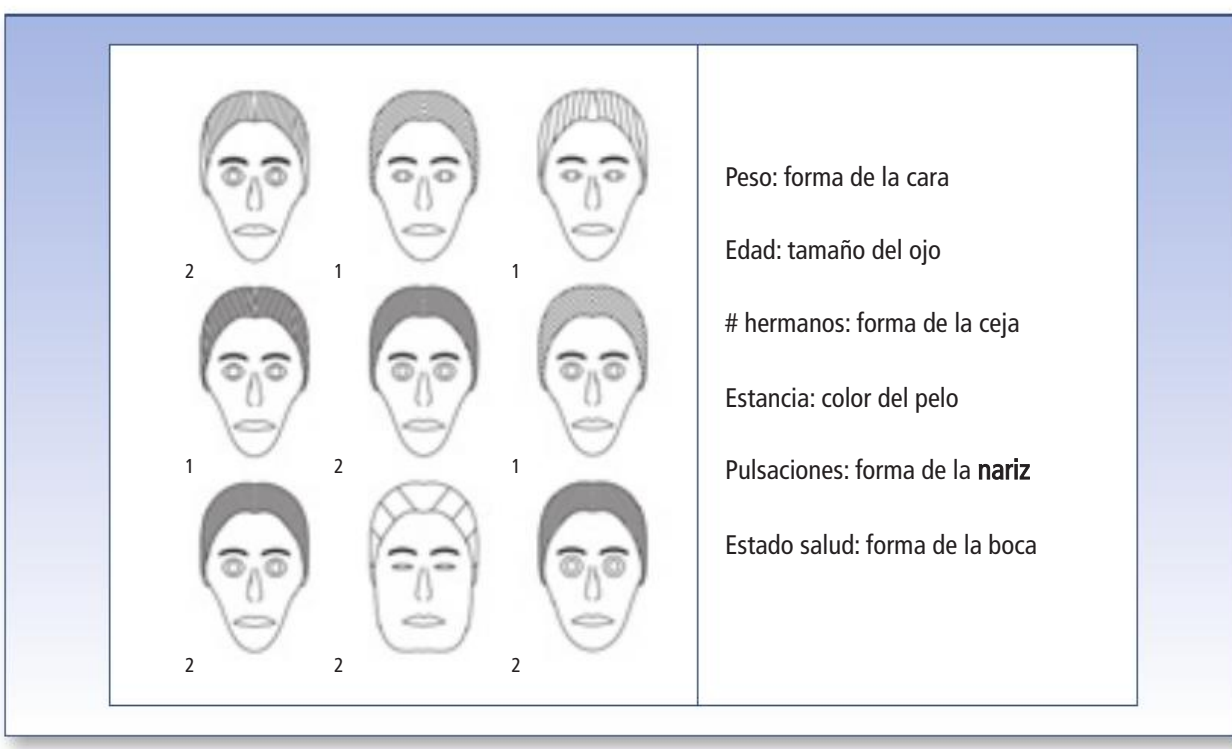


Figura 33.17 Caras de Chernoff para mostrar la información "Primeros nueve recién nacidos en un hospital en 2014".

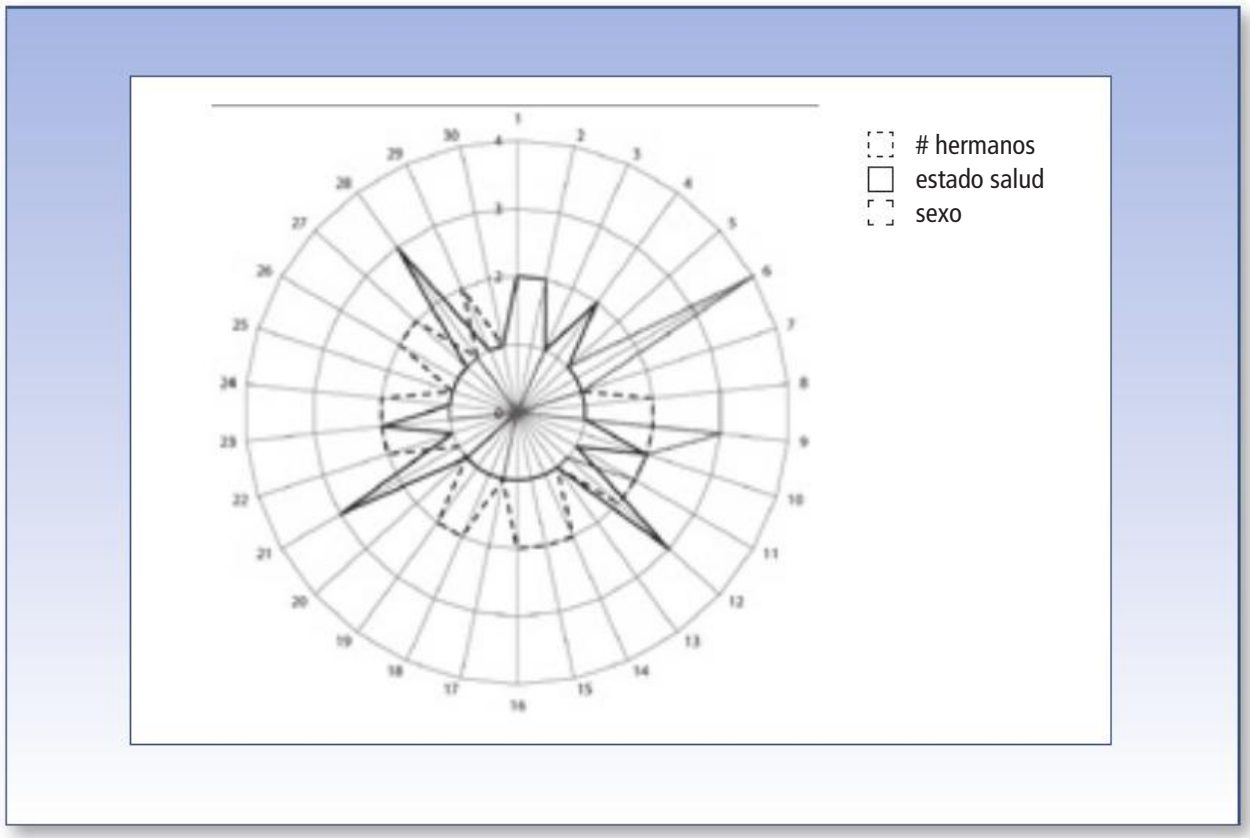


Figura 33.18 Gráfico de radar para mostrar la información "Variables sexo, estado de salud al nacer y número de hermanos".

de forma similar a como lo hace el gráfico de caras de Chernoff[21].

Las observaciones que presenten una magnitud pequeña en una variable en particular tendrán distancias al centro más pequeñas que las que presenten magnitudes más grandes para esa misma variable. La forma se logra uniendo los puntos que corresponden a cada vértice, los cuales, a su vez, dependen del aporte a los valores observados para cada individuo en cada una de las variables.

Se recomiendan para revisar si hay diferencias entre el comportamiento de los individuos, simplemente, comparando las formas de cada una de las estrellas. No se recomiendan para tamaños de muestra grandes. La **figura 33.19**, de nuevo, presenta la información de los primeros 9 recién nacidos ya mencionados.

CONSIDERACIONES FINALES

Hay una gran variedad de gráficos disponibles para la presentación estadística de datos. La recomendación general es siempre considerar, para construirlos y para usarlos, el tipo y la cantidad de variables. En la **figura 33.20** se presenta un diagrama de flujo para decidir qué tipo de gráfico utilizar según el número de variables.

También se recomienda evitar al máximo el uso de gráficos en tres dimensiones. En lo posible, no se debe duplicar información; o sea, no se debe presentar la misma información en una tabla y en un gráfico.

En la actualidad, hay una gran variedad de programas y de aplicaciones que permiten hacer, de una manera fácil, rápida y automática, construir la mayoría de los gráficos aquí presentados.

Se recomienda siempre considerar la pregunta de investigación y los objetivos del estudio para definir específicamente la información que se necesita y cómo se la va a presentar. Es importante anotar que hay, además, otros tipos de gráficos, no explorados en este capítulo, y que también se pueden utilizar y explorar durante la fase de análisis, tales como los gráficos de burbuja, los gráficos de bosque y los dendogramas, por mencionar solo algunos.

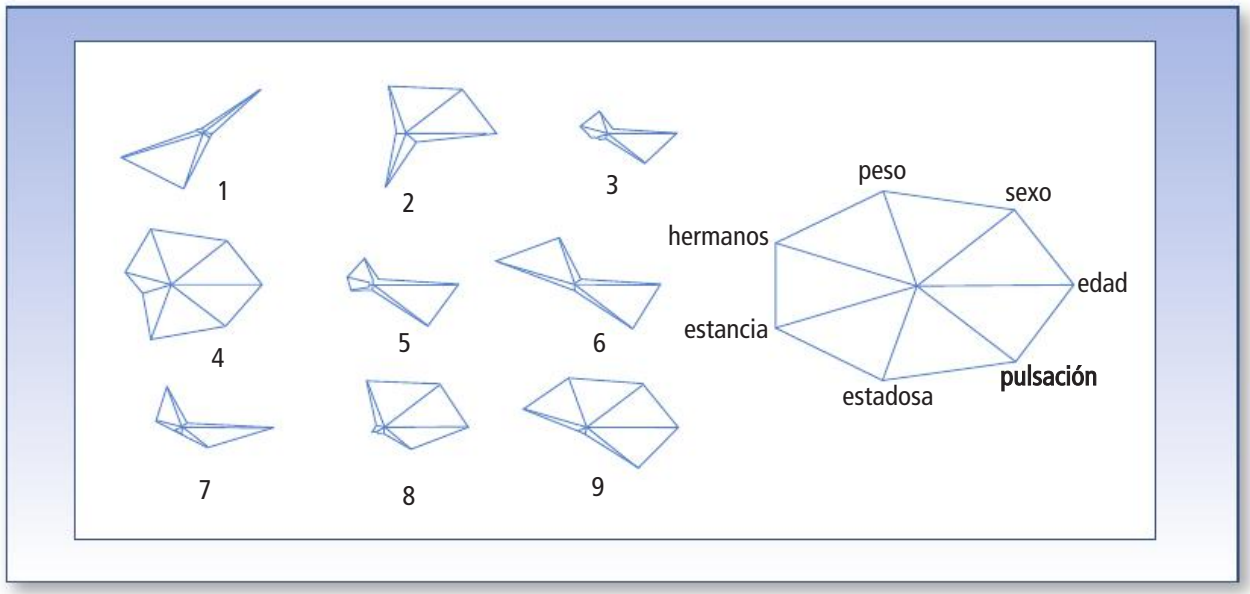


Figura 33.19 Gráfico de estrella para representar la información "Primeros nueve recién nacidos en un hospital en 2014".

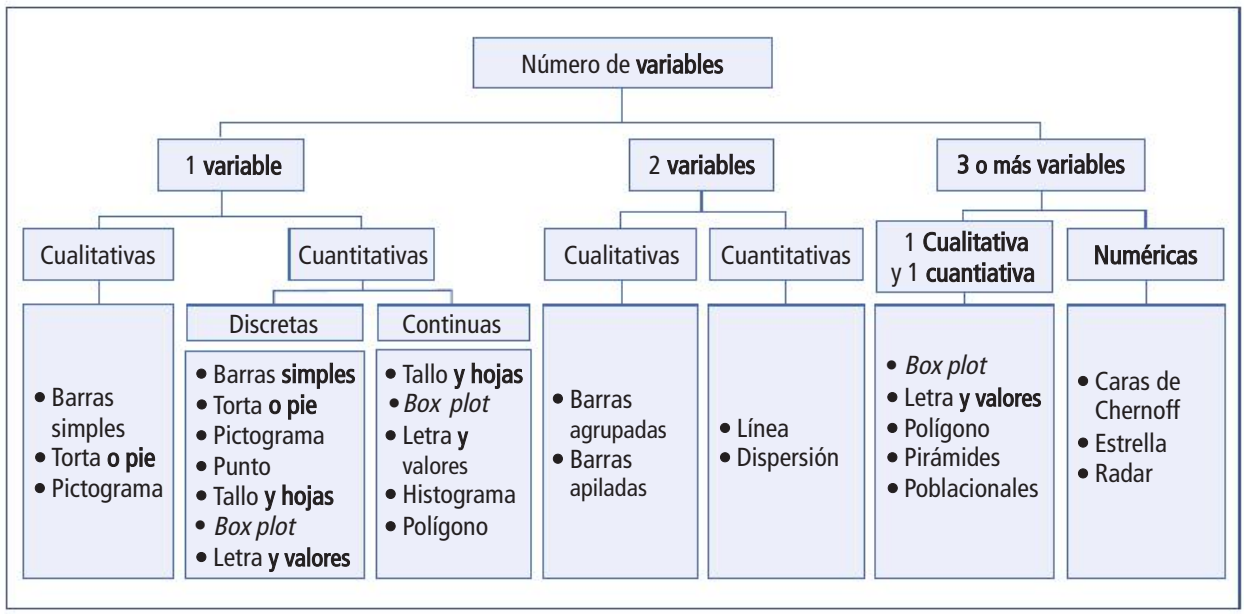


Figura 33.20 Diagrama de flujo para seleccionar el tipo de figura.

Referencias

1. Hulley SB, Cummings SR, Browner WS. Diseño de investigaciones clínicas Philadelphia: Lippincott Williams & Wilkins; 2007. http://books.google.com.co/books?id=ovFmJwAACAAJ
2. Ott R, Longnecker M. An introduction to statistical methods and data analysis Texas: Cengage Learning; 2008.
3. Namakforoosh MN. Metodología de la investigación. México: Limusa; 2000.
4. Anderson AJB. Interpreting data: A first course in statistics. Oxford, UK: Taylor & Francis; 1989.
5. Dawson-Saunders B, Trapp RG, de la Vega Sánchez ML. Bioestadística médica. México: El Manual Moderno; 2005.
6. Torre GL. Applied epidemiology and biostatistics. . Torino: SEEd; 2010.
7. Ambrosius WT. Topics in biostatistics. Nueva Jersey: Humana Press; 2007.

8. Sturges HA. The choice of a class interval. J Am Stat Assoc. 1926;21:65-6.
9. Hyndman RJ. The problem with Sturges' rule for constructing histograms. Monash Univ. 1995;1-2.
10. Doane DP. Aesthetic frequency classifications. Am Stat. 1976;30:181-3.
11. Daniel WW. Bioestadística: Base para el análisis de las ciencias de la salud. México: Editorial Limusa; 2002.
12. Hanna ARG, Rao C, Athanasiou T. Graphs in statistical analysis. En: Athanasiou T, Debas H, Darzi A, editors. Key topics in surgical research and methodology. Springer; 2010. p. 441–75.
13. Mazza R. Introduction to information visualization. Londres: Springer; 2009.
14. Norman GR, De Burgh Norman JE, Streiner DL. Bioestadística. Madrid: Harcourt; 1996.
15. Spencer NH. Essentials of multivariate data analysis. Nueva Jersey: Taylor & Francis; 2013.
16. Hofmann H, Kafadar K, Wickham H. Letter-value plots: Boxplots for large data. Am Stat. 2011;1-22.
17. Daniel WW, Nabel M, Rothman PS. Biostatistics: A foundation for analysis in the health sciences. 9th ed.. John Wiley & Sons Canada, Limited; 2009.
18. MacFarland TW. Introduction to data analysis and graphical presentation in biostatistics with R: Statistics in the Large. Springer International Publishing; 2013.
19. Colombia, DANE. Proyecciones nacionales y departamentales de población 2005-2020 [internet]. 2010 [citado 2015 ene. 15]. Disponible en http://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/7Proyecciones_poblacion.pdf
20. Iversen GR, Gergen M, Gergen MM. Statistics: The conceptual approach Nueva York: Springer; 1997.
21. Jacoby WG. Statistical graphics for visualizing multivariate data. Philadelphia: SAGE Publications; 1998.

CAPÍTULO

34

Medicina basada en evidencia

Álvaro J. Ruiz Morales

El investigador inglés Archie Cochrane escribió en 1972 que la medicina había fallado al no haber organizado resúmenes críticos, que se actualizaran periódicamente, de la información sólida proveniente de experimentos clínicos aleatorizados.

A. L. Cochrane
Londres, 1972

INTRODUCCIÓN

Tradicionalmente, el ejercicio de la medicina se ha basado en una mezcla de lógica fisiopatológica, experiencia clínica e intuición. La evaluación de este ejercicio muestra éxitos indudables, descubrimientos, avances y progresos acompañados de desarrollo de nuevas herramientas diagnósticas y de nuevas intervenciones. Parecería que el ejercicio ha seguido un patrón adecuado.

El ejercicio de la medicina tenía como base la creencia de que la calidad de la educación y la experiencia profesional garantizaban la efectividad de las intervenciones propuestas. Sin embargo, algunos estudios durante los años setenta del siglo **xx mostraron que menos del 15 %** de las intervenciones habituales se basaban en estudios sólidos de investigación, y que muchos de los tratamientos considerados efectivos no tenían buenos resultados cuando se los probaba en experimentos clínicos[1].

La educación médica estuvo centrada por muchos años en la transmisión de experiencias y de conocimientos, a veces empíricos, sin un verdadero énfasis en la necesidad de evaluación rigurosa de lo ejercido. Poco a poco, se empezó a ver la necesidad de evaluar la efectividad de intervenciones largamente utilizadas sin sustento, y que ahora han caído en desuso. Desde finales del siglo **xix y hasta casi mediados del siglo xx,** por ejemplo, se utilizó el aceite de chalmugra para el tratamiento de la lepra, hasta cuando los estudios mostraron que no tenía ningún efecto[2]; por otra parte, durante casi diez años se hizo profilaxis rutinaria antiarrítmica en los pacientes con infartos cardíacos, hasta cuando se demostró que aumentaba la mortalidad[3]; varios estudios observacionales parecían confirmar lo que sugería la fisiopatología: que la administración de terapia de reemplazo hormonal a las mujeres menopáusicas reduciría el riesgo de enfermedad coronaria, pero los experimentos clínicos a partir de 1998 mostraron que no solo no había reducción en el riesgo, sino que aumentaba el riesgo de otras condiciones, como las enfermedades tromboembólicas y de la vía biliar, además de algunas formas de cáncer[4].

Se hizo evidente que cada uno de los elementos que componían el camino largamente recorrido de la transmisión de la experiencia clínica, aparentemente adecuado, tiene debili-

dades y problemas potenciales. La intuición es, por supuesto, la más débil de las tres, por cuanto sus bases no son tangibles, ni sus resultados son reproducibles ni fácilmente evaluables. La intuición puede ser el resultado de la automatización de un conocimiento, o el desarrollo de destrezas como resultado de repetir un patrón que se observa o se produce. La lógica fisiopatológica tiene limitantes, pues no siempre lo que es conceptualmente lógico se traduce en un concepto aplicable.

EJEMPLO

Un hombre de 63 años, hipertenso crónico y con insuficiencia cardíaca consulta porque, si bien recibe antihipertensivos y diuréticos, sigue con disnea de esfuerzo. En el examen no hay edemas ni ingurgitación yugular, y el ecocardiograma muestra muy disminuida la fracción de eyección: al 10%. Se le inició y se le tituló un bloqueador beta hasta lograr el control de los síntomas.

Pero los bloqueadores beta disminuyen la fuerza de contracción y la frecuencia cardíaca. Tradicionalmente se los contraindicó en la insuficiencia cardíaca, dada la fisiopatología de la enfermedad. Pero por investigación sólida se llegó a indicarlos formalmente en los estados más avanzados de la insuficiencia cardíaca, ya que disminuyen la actividad, muy aumentada, de los sistemas nervioso central y del sistema renina-angiotensina, y mejoran la eficiencia de la función miocárdica[5]. *Lo que la fisiopatología y el conocimiento iniciales contraindicaban se convierte en indicación formal por los resultados de la investigación*.

La experiencia clínica es un aliado valioso del clínico, pero debe utilizarse con precaución, pues no se basa en una recolección organizada, sistematizada y completa de intervenciones. Por el contrario, hay un alto componente subjetivo, además de altas probabilidades de sesgo en el recuerdo y de cambio de las condiciones externas en las que se ejerce. El riesgo de memoria selectiva es muy alto[6], ya que el clínico tenderá a recordar mejor sus casos buenos, o los muy malos, y, posiblemente, recordará con menos facilidad los que tuvieron complicaciones o significaron fracasos de las intervenciones. Y si los recuerda, tenderá a considerarlos casos excepcionales; sobre todo, si encontró alguna razón o una explicación potencial para las complicaciones.

Como resultado de las debilidades de los componentes habituales de la práctica, es ideal que el clínico tenga una ayuda formal y sólida en su ejercicio, que complemente su entrenamiento médico y apoye su juicio clínico y su sentido común. Idealmente, este apoyo debe provenir de fuentes sólidas, que ofrezcan información de buena calidad y tengan amplia aceptación en la comunidad científica internacional. Dichas fuentes corresponden a la investigación clínica de óptima calidad[7].

En 1991, David Sackett y el grupo de la Universidad de MacMaster, en Canadá, propusieron una estrategia para el ejercicio de la medicina, a la cual llamaron medicina basada en evidencia (MBE)[8,9], como la integración de tres componentes: 1) evidencia proveniente de una investigación clínica sólida y de la mejor calidad; 2) la experiencia médica, y, finalmente, 3) los derechos y las preferencias del paciente.

Al proponer los resultados de investigación como componente fundamental de la práctica clínica, se hace énfasis en la necesidad de su uso, en la importancia de su evaluación y en lo inapropiado que resultaría ofrecer cuidado médico a un paciente únicamente con bases como la experiencia, la autoridad o la lógica fisiopatológica, que son insuficientes y deben ser moduladas y matizadas por los resultados de investigación de óptima calidad.

Pese a lo anterior, sería muy poco práctico proponer que se ejerza medicina exclusivamente con base en resultados de experimentos clínicos de buena calidad, pues no siempre los habrá, y hay áreas en las que no es tan fácil encontrar investigación experimental (áreas quirúrgicas, principalmente).

La principal propuesta de la MBE es que se busque siempre la mejor evidencia disponible, y cuando haya información sólida, se la utilice; pero también propone la MBE que si no la hay, se usen otras opciones como la fisiopatología, la experiencia clínica o el consenso de expertos. Pero (y esta tercera parte de la definición es fundamental) se debe tener claro que las decisiones con estas bases no pueden tener el mismo peso que las basadas en evidencia.

La propuesta de buscar el respaldo científico de investigación sólida para las decisiones médicas ha tenido una amplia aceptación, y la MBE tuvo un pico de popularidad inicial, que cedió un poco cuando empezaron a aparecer las limitaciones en su aplicación y las críticas

de quienes no estaban del todo de acuerdo con ella[10]. El proceso ha sido comparado con una adolescencia difícil, pues en sus inicios la MBE fue bien recibida, pero con el tiempo han empezado a aparecer posiciones que la controvierten (al menos, parcialmente).

El concepto de MBE ha evolucionado, y poco a poco ha madurado a un enfoque que busca dar valor a las evidencias, no siempre provenientes de experimentos clínicos aleatorizados (ECA) y con doble enmascaramiento; que, además, busca dar peso e importancia a la experiencia clínica (ojalá, sistematizada); que refuerza la importancia de la individualización del paciente y la de tener en cuenta sus circunstancias completas, y que busca reforzar la metodología de evaluación de áreas diferentes de la terapéutica, como las pruebas diagnósticas y los estudios de riesgo y pronóstico.

MBE: LO QUE ES Y LO QUE NO ES

La MBE tiene como base dos principios fundamentales[11]. El primero, la fuerza de la información tiene una estructura jerárquica, en la que es fácil identificar los dos extremos: a) información proveniente de estudios clínicos sólidos, de metodología impecable y con amplia validez externa; y b) información originada en observaciones empíricas, no estandarizadas, o con muy pocas observaciones. En el medio hay una amplia variedad de orígenes, que ofrecen solidez variable, a veces clara y fácil de identificar. El segundo principio fundamental es que la evidencia proveniente solo de la investigación nunca es suficiente para la toma de decisiones clínicas, pues siempre debe matizársela con el análisis de los beneficios y de los riesgos potenciales y con los costos, las dificultades y los inconvenientes[12].

La MBE puede ayudar al clínico a solucionar (al menos, parcialmente) muchos de los problemas diarios: ayuda a disminuir (o, al menos, a cuantificar) la incertidumbre ante la necesidad de uso de una intervención de alto costo, con riesgos inherentes o con dificultades logísticas; puede mejorar el proceso de toma de decisiones, que tradicionalmente se hace con bases débiles (o, al menos, puede hacer explícitas las debilidades del proceso); apoya al clínico a quien le falte experiencia con una recomendación de bases sólidas; puede ayudar a racionalizar las decisiones en casos donde la información proviene de experiencia no sistematizada; ofrece soluciones prácticas y rápidas a quienes necesitan información urgente, pero de buena calidad; ofrece estrategias sólidas para enfrentar la avalancha de nueva información a la que se enfrenta.

Finalmente, la MBE le ofrece al clínico herramientas prácticas de análisis que le permitan manejar la presión de la industria farmacéutica con medicamentos o intervenciones nuevas, y le ofrece respuestas científicas a las inquietudes de sus pacientes y los familiares de estos; sobre todo, en la época actual, de grandes facilidades informáticas y de muy amplia disponibilidad de información en Internet y en bases de datos electrónicas.

ELEMENTOS PARA LA TOMA DE DECISIONES

En el proceso de toma de decisiones, el profesional de la salud utiliza varias herramientas y fuentes de información; casi siempre, de manera intuitiva[13], y recibidas como ejemplo o transferidas pasivamente a partir de la forma tradicional de enseñanza de las ciencias de la salud[14]. Para que el proceso dé los mejores resultados, debería estandarizarse el uso de algunas de esas herramientas, hacerse explícito el de otras y reconocerse las ventajas y los beneficios de unas y las limitaciones de otras. Las utilizadas más a menudo son:

Experiencia clínica individual

A lo largo de su ejercicio, el profesional acumula experiencia y adquiere habilidades que provienen de la repetición de situaciones en la atención médica[15]. Se trata de una invaluable fuente de información, que tiene, sin embargo, una seria limitación: como la recolección de experiencias no es sistemática, puede ocurrir que el recuerdo esté sesgado o sea incompleto[16]. Y como no tiene contraste (vale decir, un grupo control) no es un método adecuado para evaluar la bondad ni los riesgos de la intervención. Aun si todas las experiencias son buenas, no hay forma de saber si podrían haber sido mejores los resultados con otra intervención. La experiencia clínica merece consideración y respeto, y debe valorársela; pero,

al mismo tiempo, debe reconocerse que tiene limitaciones, y que no puede ser la única fuente de conocimiento para la toma de decisiones.

El uso de la experiencia individual tiene otra limitación que debe conocerse: a medida que se acumulan casos similares, se tiende a reemplazar el conocimiento por la habilidad. En los tres dominios conocidos del aprendizaje[17] (conocimientos, destrezas y actitudes), se hallan sintetizados los aspectos necesarios para la toma de decisiones. Se tienen conocimientos que permiten saber las causas y los efectos, sus interacciones y los aspectos externos que pueden influirlas. En la actitud cuentan el interés por el conocimiento y por la solución de problemas, así como la curiosidad intelectual y la disposición para encontrar soluciones a problemas a los que se da importancia. Finalmente, la habilidad o la destreza permiten reconocer patrones, aplicar soluciones, evaluar intervenciones y ejercitar acciones que pueden tener como base, o no, el conocimiento.

Quiere decir todo lo anterior que se puede reconocer adecuadamente un patrón (como un dolor precordial opresivo acompañado de disnea), identificar correctamente algunas características clínicas (para el caso, palidez y cianosis periférica, hipotensión y actitud de ansiedad extrema) y proponer un diagnóstico preciso (por ejemplo, infarto del miocardio con falla ventricular izquierda). Más aún, se puede instaurar el tratamiento adecuado (trombólisis, analgesia, soporte inotrópico, etc.); todo eso, sin tener el conocimiento fisiopatológico de las causas del dolor, del motivo para la aparición del síndrome coronario agudo, y sin saber la razón por la cual se usa o no oxígeno, o por la que se toma la decisión de trombólisis, y no la de angioplastia (o viceversa). Se trata de aplicar experiencias repetidas que dejaron huella como destrezas, pero no están respaldadas por conocimientos.

Información proveniente de investigación

La investigación es la más sólida de las fuentes de información, pero también puede ser la más variable en calidad, por cuanto los estudios que la originan pueden ser series de casos muy sesgadas, ECA sólidos, metaanálisis inapropiados, estudios de casos y controles impecables, experimentos llenos de errores, etc.

Solo si se tiene entrenamiento formal en la apreciación crítica de la literatura se podrá evaluar correctamente la validez interna (cuán correcto es el estudio desde el punto de vista metodológico y científico), se tendrá capacidad para interpretar adecuadamente los resultados[18] (valor de los intervalos de confianza [IC], verdadero valor de un resultado específico de ***reducción relativa de riesgo*** (RRR), etc.); y, finalmente, solo con conocimientos amplios en el tema, en el entorno y en la metodología se podrá hacer la evaluación final, y, más importante, se podrá calificar la ***validez externa del estudio***; o sea, su capacidad para ser extrapolada a la población de interés[19]. En esta evaluación se tienen en cuenta no solo los criterios de admisión, sino las condiciones particulares de la enfermedad en estudio, y las de la población en relación con la condición de interés y con la intervención, entre otros aspectos.

Expectativas y deseos del paciente y de su entorno

No basta con tener información de buena calidad (***evidencia***) que apoye una intervención en la que se tiene experiencia y destreza. Es importante conocer los deseos y las expectativas del paciente a quien serán aplicadas, saber cuáles son sus condiciones y sus particularidades que podrían hacer por completo inapropiada una intervención calificada como perfecta[20].

Consideraciones locales logísticas, económicas, legales y éticas

Finalmente, las consideraciones generales en relación con la intervención deben estar enmarcadas dentro de los aspectos económicos y logísticos (particularmente, en nuestros sistemas de salud), así como en los aspectos legales y éticos que podrían hacer una intervención ilógica, o difícil de aplicar, o desproporcionada con el entorno, o imposible por limitaciones técnicas o de recursos financieros o humanos.

En general, las mencionadas fuentes de información para la toma de decisiones tienen la misma importancia por sí solas, y no podría haber una jerarquía o un orden específico al respecto. Todas y cada una deben ser consideradas en su momento particular. Lo ideal sería que se pudiesen armonizar y tener en cuenta globalmente.

Este es, justamente, el proceso que propone la MBE: armonizar e integrar todos esos componentes, para que se dé valor a la experiencia, pero siempre a la luz de la evidencia (es decir, de información irrefutable proveniente de la investigación) y enmarcada dentro de las particularidades del sujeto a quien se aplicarán y de las condiciones y las circunstancias que lo rodean. No hay ninguna intención en la propuesta de MBE de quitar importancia a alguno de estos aspectos. Se trata de darles el valor adecuado y proporcionado. No se busca eliminar la experiencia personal ni demeritarla, como tampoco dar una preponderancia irrestricta a la literatura científica de buena calidad.

Para la toma de decisiones en medicina se necesita, de manera indispensable, información. Y la ***información*** debe ser sólida. Si bien los procesos de ***búsqueda de información*** y la posterior ***clasificación según su relevancia*** son muy importantes, el paso realmente crítico es la ***evaluación crítica*** de la información encontrada.

A menudo, la medicina clínica se hace con bases curiosas: una de las más comunes es hacer *medicina basada en apariencias*, pues, a veces, parece que un nuevo examen sirve, o parece ser que una droga ha mostrado ser útil. Sin embargo, la base sólida no existe. Tampoco hay mucho sustento en quienes ejercen ***medicina basada en conferencias***: información recibida de expertos que a veces sí lo son, y a veces, no tanto. A menudo, impactan la facilidad de palabra o la capacidad para convencer (***medicina basada en grandilocuencia o en vehemencia***), aunque no haya fundamento sólido en lo que se expone. En situaciones clínicas serias, ante la necesidad de decisión urgente, se hace ***medicina basada en emergencias***, que soluciona problemas a través de lo primero que se ofrece o de lo más fácil de hacer.

Otras veces, la presión por usar lo más corriente favorece la selección de intervenciones que se muestran por estar de moda (***medicina basada en efervescencia***) o lo que recomienda el director del servicio o el profesor más importante (***medicina basada en eminencia***). También, en muchos casos, se insiste en que los casos que ha acumulado a través de los años le dan argumentos al clínico para proponer su plan como el mejor o como probado, aunque solo lo haya sido a través de casos no sistemáticamente recolectados (***medicina basada en experiencia***).

En todos los casos anteriores, el común denominador es la falta de elementos sólidos en la toma de decisiones. Igual sucede cuando se ejerce medicina defensiva para evitar demandas (***medicina basada en jurisprudencia***) o cuando se busca, principalmente, el beneficio económico por encima del bienestar del paciente (***medicina basada en gerencia***) o cuando se hace lo que recomienda quién más fuerte habla o quien cede menos o quien aparece como el más importante (***medicina basada en prepotencia***).

Idealmente, la toma de decisiones debe tomar en cuenta todos los valiosos elementos mencionados previamente, pero que no son suficientes por separado para tomarlos como base de una intervención. La experiencia es importante, pero casi nunca se la evalúa sistemáticamente; la importancia de un distinguido profesor de gran experiencia es indudable, pero existen los riesgos de sesgos; debe contemplarse el aspecto económico en las decisiones, pero como un condimento de la efectividad y de la seguridad, y nunca como el ingrediente principal en el análisis de decisiones[21,22].

Ahora bien, si se quiere identificar el aspecto más importante en la toma de decisiones, sin lugar a dudas, ese es demostrar la utilidad y la seguridad de una intervención a través de investigación de la mejor calidad. Matizada por la experiencia, modulada por las necesidades y los deseos del paciente, pero como base del proceso diagnóstico. A esta modalidad de medicina, que busca tener un sustento incontrovertible en cada decisión y exige la calificación de la solidez de las bases de las decisiones se la ha propuesto como la base de la MBE.

PROCESO DE LA TOMA DE DECISIONES EN MEDICINA CLÍNICA

El proceso de MBE es un proceso explícito que traslada la evidencia de la investigación al entorno clínico, al pie de la cama del paciente. El proceso incluye varios pasos concretos:

1. El clínico necesita tomar una decisión para la cual requiere información.
2. Se traduce la necesidad de información en una pregunta concisa.
3. Se busca sistemáticamente y de manera completa la información disponible en la literatura.

4. Se evalúa la calidad de la información a través del análisis crítico de la literatura.
5. Se aplican los resultados a la clínica.
6. Finalmente, se hace una revisión de los resultados clínicos, para retroalimentar el proceso.

Si bien el proceso se parece mucho al proceso habitual de la práctica clínica, tiene con él algunas diferencias importantes: en la clínica no siempre se hace una búsqueda sistemática de la información, y que pretenda ser completa. No siempre se hace la evaluación crítica de la información encontrada. Y no siempre se hace la retroalimentación del proceso a través de la observación de los resultados obtenidos luego de la toma de decisiones.

La descripción del proceso lo muestra sólido. Pero debe reconocerse que la fase de búsqueda de la literatura (si es exhaustiva y sistemática), para una posterior evaluación crítica, que necesita conocimiento de metodología de investigación, habilidades en la lectura crítica y bases bioestadísticas, no siempre es un proceso factible en la vida real, en el diario trajinar de la medicina clínica.

La importancia del proceso y el potencial resultado del uso racional de la evidencia hacen que se deba dar prioridad al ejercicio de la **MBE.** Pero si bien debe buscarse que todo clínico sea independiente en su capacidad para hacer búsquedas y para hacer una apreciación crítica de la evidencia, también debe pensarse que hay caminos más prácticos para encontrar las respuestas a las preguntas habituales, a través de lo publicado por personas que ya hicieron el trabajo.

El énfasis debería estar, más bien, en ***buscar,*** *en lo posible, la mejor evidencia disponible, a través de búsqueda en los sitios especializados de información ya procesada; si no hay información disponible, debe hacerse el proceso largo de búsqueda, apreciación y recomendación final*. El orden del proceso debe empezar, entonces, por las ***guías de práctica clínica (GPC) basadas en evidencia***, por los *consensos basados en evidencia* y por las publicaciones especializadas que tienen las recomendaciones, basadas en evidencia, para una mejor atención en la salud.

Algunos de los pasos propuestos merecen ser comentados.

Formulación de una pregunta

En el **capítulo 2** "La pregunta de investigación", se mencionó el proceso de formulación de una pregunta y de una hipótesis, sus condiciones y sus características mínimas, y cómo el tipo de pregunta depende de la utilidad que se le va a dar a la información. Una pregunta que tenga claros los componentes mínimos necesarios podrá orientar al clínico en la búsqueda y la posterior aplicación *evaluable* de los resultados del proceso. La pregunta, cabe recordarlo, debe tener claridad en la *población* de estudio, estar definida de manera inequívoca y corresponder con el sujeto o el grupo de sujetos a quienes se aplicarán los resultados; debe, además, tener una definición clara, precisa y que permita reproducir la ***intervención que se evalúa***, así como la ***intervención de control***; finalmente, debe haber claridad en los ***desenlaces*** con los cuales se evaluarán los resultados.

Búsqueda de literatura

Este paso era importante, pero impensable como obligatorio hasta hace muy pocos años. Las razones fundamentales para ello eran la escasa disponibilidad de literatura en el medio habitual clínico y las dificultades para hacer búsquedas con los aparatosos, no siempre presentes y casi nunca completos tomos del *Index Medicus*, que producían más de un dolor de cabeza por no estar disponibles, por la falta de las hojas que se necesitaban o porque ofrecían referencias difíciles de conseguir para solucionar el problema.

En el momento, hay dos elementos de importancia que han permitido resolver el problema y han permitido la disponibilidad inmediata, o muy fácil y casi ilimitada, de información para el médico clínico en un ambiente convencional:

a. **Internet, los computadores y las redes de información:** Hoy día, se pueden hacer búsquedas gratuitas en tiempos mínimos, a través de una conexión a la red mundial, en muchos sitios. Las conexiones están, cada vez más a menudo, disponibles en bibliotecas, hospitales, oficinas y hogares, así como en sitios públicos (cafés, restaurantes, aeropuertos y salas de espera).
b. **Bases de datos específicas de información ya recolectada y procesada:** No siempre es práctico hacer una búsqueda de la literatura cuando se va a tomar una decisión. Pero siempre es

práctico y adecuado utilizar la búsqueda y la evaluación que alguien con experiencia hizo y puso a disposición en una colección de recomendaciones. Las ***GPC*, las *recomendaciones y consensos hechos con base en evidencia* y las colecciones de investigación** y de metaanálisis, como la ***Biblioteca Cochrane* o las publicaciones periódicas** de información sólida y sistemáticamente recolectada, como ***Evidencia Clínica***, permiten el ejercicio práctico y adecuado de la medicina: se puede buscar cuál es el soporte en investigación de buena calidad para una intervención, y se puede, en poco tiempo, calificar la racionalidad de una decisión según el sustento que pueda encontrarse.

Evaluación de la información obtenida

Se busca analizar la calidad de la información a través de preguntas como: ¿cuál es la validez del estudio? ¿Cuál es la generalizabilidad de los resultados? ¿Cuáles son los resultados, y cómo se interpretan a la luz del concepto de significación clínica?

La primera pregunta, sobre ***validez interna*** del estudio, busca evaluar si el resultado puede ser tomado como cierto o si puede deberse a factores externos. Cabe mencionar que el proceso para llegar a la aceptación de un resultado debe seguir los siguientes pasos:

1. **El resultado puede deberse a un sesgo:** El sesgo se refiere a una desviación sistemática de la información, siempre en el mismo sentido.

EJEMPLO

Si se hace un estudio en madres de niños con malformaciones congénitas para preguntarles por medicamentos tomados durante el embarazo, habrá un número significativamente mayor de madres que recuerden haberse expuesto a fármacos, en comparación con las madres de niños normales. La diferencia puede estar dada porque la madre de un niño anormal tiene más razones para recordar, ha tenido más oportunidades para revisar lo que pasó durante su embarazo y se le ha preguntado más por parte de quienes buscan explicar la malformación del niño, si recibió medicamentos. En dicho estudio hay una asimetría: las madres de niños anormales tienen más estímulos para recordar, y, probablemente, los resultados así lo reflejen. Es difícil en tales casos asumir que la diferencia es real.

Otros ejemplos de sesgos incluyen la asignación, por parte del investigador, de los pacientes para un estudio de un medicamento nuevo, en el cual él confía: puede asignar a quienes se hallen en mejores condiciones o a quienes, probablemente, sigan con más juicio el tratamiento al medicamento nuevo de estudio, mientras que asignará al tratamiento convencional a quienes tengan más riesgos, más probabilidad de complicaciones, más dificultades en la adherencia, etc.

2. **El resultado puede deberse al azar:** Más o menos a menudo, los resultados son similares, y pueden deberse a efecto del azar.

EJEMPLO

Si se comparan dos tratamientos, A y B, para una infección, y uno de ellos (A) controla adecuadamente la infección en el 71 % de los pacientes y el otro (B) solo controla al 68 % de los pacientes, no puede hablarse de una verdadera diferencia. Si se hubiera repetido el estudio podría haberse obtenido el mismo resultado en números, pero invertido: la intervención (A) mejoró al 68 % de los pacientes, y la (B), al 71 %. O puede obtenerse un resultado cercano, como el 70 % y el 69 %, respectivamente, o el 73 % y el 71 %, respectivamente, etc. Debe evaluarse la posibilidad de que el resultado encontrado pueda deberse al azar.

Para este propósito, se hacen pruebas estadísticas que ofrecen un resultado; en algunos casos, el conocido como ***valor de p*, cuya** interpretación es exactamente la probabilidad de que el resultado pueda deberse al azar. Si el estudio no tiene un análisis adecuado, o si se utilizó la prueba incorrecta o si la significación estadística no está dentro de los límites establecidos por la comunidad científica internacional, puede llegarse a conclusiones falsas. Más importante aún, si no se consideran los intervalos de confianza, descritos en el **capítulo 9 "Medidas de frecuencia,** de asociación y de impacto", los resultados pueden ser interpretados erróneamente.

3. **El desenlace puede ser el resultado de otros factores intervinientes, que generan confusión:** La presencia de factores diferentes de los factores evaluados que pueden influir el resultado debe ser siempre tenida en cuenta. No solo el colesterol aumenta el riesgo de sufrir un infarto del miocardio. Dicho riesgo también

está relacionado con la edad, con el sexo y con otros factores.

EJEMPLO

En un estudio reciente sobre la salud en un área urbana contigua a un aeropuerto, se encontró que quienes vivían cerca a dicho terminal aéreo tenían más altas tasas de enfermedades, de suicidios y de cáncer que quienes vivían en otras regiones de la misma ciudad. La conclusión de que vivir cerca de los aeropuertos está asociada a una mayor morbilidad es, sin embargo, falsa, pues la verdadera asociación es al estado socioeconómico: se enferma más quien es pobre, y vive cerca de los aeropuertos quien debe aceptar la opción más económica de vivienda: y entre las principales razones económicas está vivir cerca de donde hay ruido, cerca de los aeropuertos. Vivir cerca de los aeropuertos es un factor de confusión, asociado a ser pobre e indirectamente asociado a enfermarse más, pero a través de ser pobre.

Otro ejemplo claro es la falsa asociación que podría aparecer entre el consumo de café y el cáncer de pulmón. La verdadera asociación con el cáncer la tiene el cigarrillo, pero este, a su vez, se halla asociado al consumo de café.

4. **Al final, y solo después de haber descartado sesgos, el azar y los factores de confusión como responsables del desenlace, se puede aceptar la relación entre éste y la intervención que se estudió:** La segunda pregunta, sobre la generalizabilidad de los resultados, se refiere a si los datos obtenidos en la investigación pueden aplicarse a nuestra población. Los ***criterios de inclusión*** *y de exclusión* que usa el investigador para decidir cuáles pacientes pueden entrar a su estudio son los que le sirven al clínico para decidir si pueden aplicar o no los resultados en su sitio de trabajo. Si el paciente podría haber sido elegido para el estudio, se considera que puede también beneficiarse de los resultados.

Finalmente, se evalúa cuáles fueron los resultados y cuál su precisión, cuál es la significación estadística, y cuál, la significación clínica. Un resultado estadísticamente significativo puede no tener aplicación clínica porque el resultado es muy pequeño o de poco impacto clínico, o bien, porque los riesgos sobrepasan los beneficios, aunque estos sean claros, o porque los costos hacen inadecuado buscar el potencial efecto positivo.

En el proceso de MBE, el aspecto más importante es la consideración de que debe buscarse el sustento de mejor calidad, la información que, incontrovertiblemente, pueda considerarse cierta: a eso se llama ***evidencia***. Si bien no siempre existe o no siempre se la puede encontrar, es adecuada y recomendable la intención de ejercer siempre con la mejor base disponible. Y la mejor base disponible la proporciona la investigación de buena calidad.

Traducción a la práctica clínica

La transferencia del conocimiento a la práctica implica un proceso activo después del proceso de evaluación, y que tiene como lineamientos la clasificación que se propuso de la solidez de la información. Sea cual sea el método empleado, es útil tener una clasificación que permita hacer explícitos los resultados del proceso de búsqueda y evaluación.

Evaluación de resultados luego de la aplicación

Probablemente, este sea el paso más útil para el acervo global de conocimientos, y, probablemente, el menos aplicado. Una vez se toma la decisión individual y se la aplica al paciente, debería tenerse en cuenta (***actitud***) que se trató de un proceso que debe evaluarse para que los resultados retroalimenten los futuros procesos.

Críticas a la MBE

A lo largo del capítulo se han comentado los aspectos que se le han criticado al proceso de MBE. Se debe reconocer, como primera medida, que, si bien hay información útil que sugiere un mejor desempeño cuando se la aplica, no hay aún datos sólidos que permitan establecer que, en condiciones iguales, tiene mejor desempeño el médico que aplica el proceso o tiene mejores resultados el paciente a quien se le aplican[23].

Adicionalmente, se puede correr el riesgo de la "tiranía de la evidencia", la cual puede exagerar el valor de la investigación y quitar mérito a la experiencia colectiva o a la fisiopatología, que son complementos necesarios, y

no competidores excluyentes, en el proceso de toma de decisiones. O se privilegian los experimentos clínicos, a sabiendas de que a menudo tienen sesgos importantes desde su concepción y su diseño, en parte, por conflictos de interés económicos[24].

Hay algunos aspectos que aún se deben criticar en el proceso propuesto, como la falta de definición de la experiencia clínica, o el poco desarrollo y el poco estímulo dado por la MBE al proceso diagnóstico.

La definición de experiencia clínica como un punto débil de la MBE

Desde las primeras descripciones de Sackett, se menciona la importancia de la experiencia clínica como uno de los componentes de la MBE, supeditada a la evidencia, pero presente y con valor. Sin embargo, no hay ni definición ni precisión en torno a la experiencia, y el concepto aún no se ha delimitado.

Es claro que no puede tratarse tan solo del número de años acumulados en la práctica, y que tiene que haber un componente de calidad de ejercicio, de evaluación de resultados y de retroalimentación, así como de la posibilidad de estandarizar procesos y de actitud positiva que busque la mejoría permanente. Sir William Osler lo expresó de manera magistral: "El valor de la experiencia no estriba en ver mucho, sino en ver con sabiduría".

La evaluación del proceso diagnóstico como punto débil de la MBE

Como se ha mostrado en algunos estudios, la probabilidad de recibir el tratamiento correcto en la era de la MBE está alrededor del 80 %. Sin embargo, el proceso diagnóstico ha recibido menos atención en la literatura, y las propuestas de la MBE se refieren menos al diagnóstico, y mucho más, a la terapéutica. Y los métodos para evaluación crítica de la literatura y para la aplicación de evidencia a la práctica clínica se centran prioritariamente en ECA, y mucho menos, en pruebas diagnósticas, o en características operativas de estrategias o en pruebas para diagnóstico.

Si bien los estudios publicados en revistas especializadas en el enfoque de la MBE (como el ACP *Journal Club*) son, principalmente, acerca de intervenciones terapéuticas, y menos del 10 % lo es sobre pruebas diagnósticas, las necesidades expresadas por los médicos generales tienen una distribución bastante diferente, cercana al 40 %, tanto para terapéutica como para diagnóstico.

El temor habitual de parálisis ante la falta de evidencia es infundado, pero muy generalizado. El médico no bien familiarizado con el concepto de MBE puede pensar que solo se actúa cuando hay información sólida y proveniente de estudios impecables. Dado que sabe, por experiencia, que no siempre se dispone de tal información, y que los estudios impecables son escasos, puede temer que deberá quedarse paralizado en tales situaciones. La respuesta que ofrece la MBE es la oportunidad de cuantificar o calificar la solidez de la información disponible y actuar en consecuencia: con gran confianza cuando hay evidencia, y con menos confianza cuando no la hay, ***pero siempre actuar***. En algunos casos tendrá que basarse el clínico tan solo en experiencias previas ante la falta (real, por supuesto) de información sólida proveniente de investigación de buena calidad.

Uno de los temores más reales de quienes practican medicina es que ante cada situación de conflicto con un paciente se vean obligados a desarrollar interminables búsquedas bibliográficas, a gastar largas jornadas leyendo y criticando información y a hacer ingentes esfuerzos para clasificar la información, obtener la evidencia y producir una recomendación que, finalmente, pueda ser aplicada al paciente, quien se halla, entretanto, abandonado a su propia suerte. Si bien el proceso de búsqueda de información, crítica y producción de una conducta es la base de la MBE, debe saberse que, cada vez más a menudo, se encuentran las recomendaciones de quienes recorrieron ya el camino, y lo hicieron de manera práctica y muchas veces estructurada, en la forma, por ejemplo, de GPC basadas en evidencia.

CONCLUSIÓN

En el proceso de práctica de la medicina, es indispensable basar las decisiones en la mejor y más sólida información disponible; o sea, *en evidencia*. El método propuesto permite sistematizar la búsqueda y la evaluación de la información, y facilita evaluar las decisiones tomadas. La necesidad enorme de lectura para

actualización hace indispensables métodos de búsqueda sistematizados, comprehensivos y eficientes. Actualmente hay muchas publicaciones especializadas (***Evidence Based Medicine Journal***), Evidencia Clínica (***Clinical Evidence***), bases de datos (***The Cochrane Collaboration***), y páginas de red dedicadas (HIRU, ACP on line, BMJ) que ayudan al médico ocupado presentándole, a veces, no solo la información, sino los metaanálisis existentes o una apreciación crítica de ella, y, bastante a menudo, guías de práctica basadas en la metodología propuesta.

La toma de decisiones con bases sólidas de evidencia tiene una probabilidad de éxito mucho mayor que la basada en intuición, en la experiencia no sistematizada o en la "lógica fisiopatológica" no probada. La mejor decisión, cabe aclararlo, no siempre es la menos costosa, por lo cual el proceso de práctica con MBE no puede equipararse con reducción de costos; al menos, no con una inmediata, aunque a menudo en la evaluación a largo plazo pueda hallarse que las decisiones con bases sólidas tienen mejores eficiencias que las tomadas con bases débiles, de experiencia no sistematizada o con bajo número de observaciones. Se busca la mejor decisión, la más eficaz (operativa en condiciones ideales), la más efectiva (útil en condiciones reales) y la más eficiente (la que produce los mejores resultados con los menores costos y los menores riesgos).

Referencias

1. Wennberg JE. Tracking medicine. A researcher's quest to understand health care. Oxford: Oxford University Press; 2010.
2. Kar HK, Gupta R. Treatment of leprosy. Clin Dermatol. 2015;33:55-65.
3. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. N Engl J Med. 1991;324:781-8.
4. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA. 1998;280:605-13.
5. Yancy CW, Jessup M, Bozkurt B, et al. ACCF/AHA Guideline for the management of heart failure: executive summary: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Circulation. 2013;128:1810-52.
6. Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. BMC Health Services Research. 2003;3:14.
7. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine: how to practice and teach it, 4th ed. Edinburgh, Churchill Livingstone; 2010.
8. Evidence-based medicine. A new approach to teaching the practice of medicine. Evidence-Based Medicine Working Group. JAMA. 1992;268:2420-5.
9. Guyatt GH. Evidence-based medicine. ACP Journal Club. 1991;A-16.
10. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ. 2014;348:18-219.
11. Haynes RB, Sackett DL, Gray JM, et al. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. ACP J Club. 1996;125:A14-6.
12. Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. Lancet. 1997;349:126-8.
13. Davis DA, Thomson MA, Oxman AD, et al. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA. 1995;274:700-5.
14. Guyatt G, Sackett D, Taylor DW, et al. Determining optimal therapy--randomized trials in individual patients. N Engl J Med. 1986;314:889-92.
15. Sackett DL, Rennie D. The science of the art of the clinical examination. JAMA. 1992;267:2650-2.
16. Greenhalgh T. Narrative based medicine: narrative based medicine in an evidence based world. BMJ. 1999;318:323-5.
17. Anderson LW, Krathwohl DR, Airasian PW, et al. A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives. New York: Pearson, Allyn & Bacon; 2001.
18. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? JAMA. 1998;280:280-2.
19. Ioannidis JP. Effect of the statistical significance of results on the time to completion and pu-

blication of randomized efficacy trials. JAMA. 1998;279:281-6.
20. Gill P, Dowell AC, Neal RD, et al. Evidence based general practice: a retrospective study of interventions in one training practice. BMJ. 1996;312:819-21.
21. O'Brien BJ, Heyland D, Richardson WS, et al. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA. 1997;277:1802-6.
22. Wasserfallen JB. Medical-economic analysis: a necessary complement to "evidence based medicine". Rev Med Suisse Romande. 2001;121:265-8.
23. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79:S70-S81.
24. Wyer P, Alves da Silva S. 'One mission accomplished, more important ones remain': commentary on Every-Palmer, S., Howick, J. (2014) How evidence-based medicine is failing due to biased trials and selective publication. Journal of Evaluation in Clinical Practice, 20 (6), 908-914. J Eval Clin Pract. 2015 Feb 26. doi: 10.1111/jep.12330. [Epub ahead of print]

CAPÍTULO

35

Apreciación crítica de la literatura

JUAN MANUEL LOZANO LEÓN

Pero si anhelas ser sabio
ojalá viéndome aprendas
a rumiar, y rumiar mucho,
cada bocado de ciencia.
Sin rumiarlo bien, no engullas
ni una línea ni una letra;
el que aprende como un loro,
loro ignorante se queda.

RAFAEL POMBO

INTRODUCCIÓN

Los versos del poeta, traductor y diplomático colombiano Rafael Pombo son parte de la respuesta de un buey a un niño que, camino a la escuela, le reclama por permanecer acostado rumiando, mientras él estudia. El consejo de no engullir sin rumiar se aplica perfectamente a quienes utilizan literatura científica.

Para tomar decisiones clínicas correctas, se requiere información adecuada, que proviene, habitualmente, de dos grandes fuentes: del paciente y de la investigación clínica. La información proveniente de los pacientes, acumulada, se constituye en la experiencia del médico. Las dos fuentes son necesarias para brindar cuidado efectivo de salud[1].

La información de los individuos se obtiene a través del interrogatorio clínico, del examen físico y de otras actividades complementarias, como las pruebas paraclínicas. Por su parte, la información acerca de la investigación clínica se obtiene de fuentes complementarias, que incluyen el contacto personal con colegas, las conferencias y reuniones clínicas, los cursos de educación continuada, el contacto con representantes de la industria farmacéutica y la lectura de libros, monografías, boletines de noticias y revistas.

Varias encuestas evidencian que la lectura de revistas es una de las actividades de educación continuada más importantes[2]. Sin embargo, los clínicos que buscan información proveniente de investigación en revistas enfrentan varios desafíos: dificultades para localizar los artículos apropiados dentro del ya enorme (y cada vez mayor) volumen de la literatura; obstáculos en el acceso a las bibliotecas y a los textos; costos (en términos tanto de tiempo como de dinero), y, algunas veces, la brecha entre el momento en el que se encuentra nueva información y la oportunidad de aplicarla[2].

Superados dichos obstáculos, aparece un desafío más: evaluar la calidad y la aplicabilidad de la información obtenida antes de emplearla para tomar decisiones clínicas. Aunque la mayoría de los artículos científicos pasa por el proceso de revisión por pares, lo que reduce el riesgo de publicaciones con defectos, no es apropiado

EJEMPLO

Ejemplo: ¿Las mujeres que tuvieron apendicitis aguda perforada durante su niñez tienen tasas de infertilidad más altas que las de las mujeres que no tuvieron la enfermedad?

¿Qué tipo de estudio se adelantó?

La **tabla 35.1** muestra los diseños "ideales" para estudios que tratan de responder cada una de las preguntas mencionadas, así como los aspectos críticos que permiten identificar los errores metodológicos en la planificación o la ejecución del estudio[1]. Esos diseños corresponden a estudios *primarios*, los cuales obtienen información a partir de sujetos humanos para responder la pregunta científica de interés.

La **tabla 35.2** describe los tipos más comunes de estudios ***integrativos***, que responden preguntas de investigación reuniendo la información disponible en otros estudios. Este grupo incluye las revisiones sistemáticas (metaanálisis) de la literatura (RSL), las guías de práctica clínica (GPC) y los análisis económicos. Pese a emplear datos publicados previamente, los estudios integrativos son también investigación original.

¿El estudio fue apropiado para la pregunta de investigación, y se adelantó correctamente?

Una vez el lector tiene clara la pregunta de investigación, con el fin de establecer si el estudio reportado en un artículo fue adecuado para responderla, deben evaluarse los métodos empleados por los autores. Existen guías de eva-

Tabla 35.1 Aspectos críticos en la planeación y la ejecución de estudios primarios para responder las cuatro preguntas más comunes en investigación clínica

Causalidad/riesgo
Diseño ideal: experimento clínico aleatorizado (ECA); más a menudo, estudios de cohorte o de casos y controles. Estudio de cohorte: aparte de la exposición de interés, ¿el grupo expuesto y el no expuesto comenzaron y terminaron con el mismo riesgo del desenlace? • ¿Los pacientes fueron similares respecto a factores pronósticos de los cuales se sabe que están asociados al desenlace (o se hizo un ajuste estadístico)? • ¿Las circunstancias y los métodos para detectar el desenlace fueron similares? • ¿El seguimiento fue suficientemente completo? • Estudio de casos y controles: ¿los casos y los controles tuvieron el mismo riesgo (probabilidad) de hallarse expuestos en el pasado? • ¿Los casos y los controles fueron similares respecto a la indicación o las circunstancias que llevaron a la exposición? • ¿Las circunstancias y los métodos para detectar la exposición fueron similares para los casos y los controles?
Diagnóstico
Diseño ideal: estudio de corte transversal que utilice la prueba de interés y un patrón de oro. • ¿Los pacientes participantes presentaban un dilema diagnóstico? • ¿Los investigadores compararon la prueba con un estándar de referencia apropiado e independiente? • ¿Quienes interpretaron la prueba y el estándar de referencia estuvieron enmascarados respecto a los otros resultados? • ¿Los investigadores aplicaron el mismo estándar de referencia a todos los pacientes, independientemente de los resultados de la prueba evaluada en la investigación?
Terapia
Diseño ideal: ECA. • ¿El grupo experimental y el de control comenzaron con el mismo pronóstico? • ¿Se mantuvo un balance en el pronóstico a lo largo del progreso del estudio? • ¿Los grupos se hallaban balanceados en pronóstico al final del estudio?
Pronóstico
Diseño ideal: cohorte concurrente. • ¿La muestra de pacientes fue representativa? • ¿Los pacientes fueron suficientemente homogéneos respecto al riesgo pronóstico? • ¿El seguimiento fue suficientemente completo? • ¿Los criterios del desenlace fueron objetivos y sin sesgo?

Fuente: Elaboración del autor.

Tabla 35.2 Aspectos críticos en la planeación y la ejecución de estudios integrativos para responder las cuatro preguntas más comunes en investigación clínica

Revisión de la literatura • ¿La revisión pretendió dar respuesta a una pregunta clínica sensible? • ¿La búsqueda de estudios relevantes fue detallada y exhaustiva? • ¿Los estudios primarios fueron de alta calidad metodológica? • ¿La evaluación de los estudios fue reproducible?
Guía de práctica clínica (GPC) • ¿Las recomendaciones consideraron todos los grupos de pacientes, las opciones de manejo y los posibles desenlaces que fueran relevantes? • ¿Hay revisiones sistemáticas de la literatura (RSL) que estimen el efecto relativo de las opciones de manejo en los desenlaces relevantes? • ¿Se especifican apropiadamente los valores y las preferencias asociadas a los desenlaces?
Análisis económico • ¿Los investigadores adoptaron un punto de vista suficientemente amplio? • ¿Se reportan por separado los resultados que sean de relevancia para los subgrupos de pacientes? • ¿Los costos se midieron de manera precisa? • ¿Los investigadores consideraron el tiempo de los costos y de las consecuencias?

Fuente: Elaboración del autor.

luación, ya mencionadas, que han sido diseñadas para que las usen personas sin experiencia como investigadores, tales como la mayoría de los lectores de artículos de revistas científicas[8-10]. El objetivo fundamental de esta evaluación es identificar errores que puedan invalidar los resultados. Se define la validez como el grado en el cual los hallazgos obtenidos al medir un fenómeno representan la verdadera dimensión de lo medido[11].

EJEMPLO

El resultado del peso de un niño en una balanza es válido si corresponde al verdadero peso del niño.

Los estudios de investigación son, en tal sentido, no muy diferentes de una balanza: solo son instrumentos diseñados para medir un fenómeno complejo. Como instrumentos de medición, son susceptibles de errores que pueden producir resultados falsos o no válidos. Las causas de errores en medición que pueden invalidar los resultados pueden agruparse en dos grandes categorías: 1) ***sesgo*** y 2) ***variación por el azar***.

De manera similar a lo descrito en el ejemplo de medir el peso de un niño, las observaciones y las medidas que se obtengan durante una investigación pueden diferir del valor real por efecto de sesgos o por el azar, lo cual haría que sus resultados puedan no ser válidos.

EJEMPLO

Para ilustrar la relación entre sesgo y azar como fuentes de error, asumamos que se pesa a un lactante cuyo verdadero peso (que, en realidad, no conoce quien lo pesa) es de 8,0 kg. Habrá sesgo si se emplea una balanza con un defecto mecánico capaz de hacer que los resultados sean sistemáticamente diferentes (por ejemplo, ½ kg siempre por encima) del verdadero valor. En este caso, siempre y cuando se emplee la balanza para pesar a este o a cualquier niño, se obtendrá una medida falsa, que difiere del valor verdadero aproximadamente en la misma magnitud y siempre en la misma dirección. Por otra parte, el papel del azar se hace evidente al pesar al mismo niño varias veces consecutivas, pero en las mismas condiciones, incluso usando una escala en buen estado. Es probable que cada medida obtenida difiera ligeramente del verdadero valor, y así arroje resultados algunas veces superiores y a veces inferiores, que difieren del valor real, pero de forma no sistemática. La diferencia entre el valor verdadero y cada una de estas observaciones, debida al azar, se llama *error aleatorio*[11]. En la vida real, obviamente, las dos fuentes de error pueden aparecer de forma simultánea.

Hay dos aspectos diferentes de la validez que deben evaluarse en un estudio: 1) la ***validez interna*** es la extensión en la cual los resultados del estudio son correctos para los sujetos que participaron en el estudio, y está determinada por la calidad del diseño, de la recolección de los datos y de su análisis. Por su parte, 2) la ***validez externa***,

también llamada ***generalizabilidad***, es la extensión en la cual los resultados de la observación en una muestra de individuos son válidos para sujetos que no participaron en el estudio. En otras palabras, la validez externa expresa la presunción de que los sujetos del estudio son comparables a otros fuera del estudio[1,8,11]. Es muy claro que para ser aplicables a otros sujetos, los resultados del estudio deben ser válidos internamente.

A pesar de su importancia, hay evidencia que muestra que los lectores no evalúan la metodología utilizada por los autores de los artículos que leen, y pasan directamente de la introducción a los resultados, y asumen así que la metodología es correcta; vale decir, ***que los resultados son válidos***.

EJEMPLO

Fitzgerald distribuyó entre un grupo de estudiantes una copia modificada de un artículo, en la cual había insertado la sección de material y métodos de un estudio sobre la hipofisectomía en retinopatía diabética de un artículo que describía algunos aspectos hematológicos de la retinopatía diabética. Solo uno de los miembros del grupo notó el cambio; todos los lectores restantes aceptaron las conclusiones del artículo falso sin haber examinado los métodos mediante los cuales se llegó a dichas conclusiones, y sin haber notado la inconsistencia entre el estudio y los métodos de otro[12]. El autor especula que "la gran cantidad de material que se presenta en los años de formación, la presión por memorizarlo… y la comodidad que representa el dogmatismo frente a la incertidumbre clínica desalientan el sano escepticismo ante la palabra escrita"[12].

Dada la alta probabilidad, ya mencionada, de tener estudios con defectos metodológicos importantes[3-7], es inapropiado asumir que los estudios, por el solo hecho de estar publicados, se adelantaron con métodos adecuados. Las series más extensas de guías de apreciación crítica de la literatura para los lectores sugieren seguir varios pasos consecutivos para evaluar la validez y la generalizabilidad de un artículo[8]:

- Primero debe determinarse si los resultados del estudio son veraces o válidos internamente. Como ya se mencionó, eso se logra revisando los métodos utilizados por los investigadores. Solo si el artículo describe métodos que han tratado apropiadamente las principales fuentes de error se justifica proceder con los pasos siguientes. En caso contrario, cuando se considera que el estudio tiene errores suficientes para invalidar los resultados, no es útil continuar su lectura.
- En segundo término se debe proceder a determinar la importancia clínica de los resultados del estudio. Para ello, se evalúa la magnitud de los resultados, su dirección y la confiabilidad (***repetitividad***) de los valores que miden los desenlaces finales.

EJEMPLO

Un ECA hipotético comparó la incidencia de infarto del miocardio en un grupo de personas expuestas a un hipolipemiante con otro grupo expuesto a placebo. El riesgo relativo (RR) en los sujetos expuestos al medicamento, al comparar con quienes recibieron placebo, fue de 0,77 (***magnitud del riesgo***); ello quiere decir que el riesgo de infarto del miocardio al recibir el medicamento se redujo en el 23 % (***dirección del riesgo***). Por último, los intervalos de confianza (IC) del RR, que fueron de 0,54-0,88, miden la confiabilidad de los resultados.

Finalmente, el lector debe establecer la medida en la que dichos resultados le serán de utilidad para brindar cuidado a sus propios pacientes, lo cual implica establecer la validez externa.

Las secciones que siguen amplían algunos de los aspectos relacionados con los mencionados tres pasos. Sin embargo, conviene hablar de dos puntos antes de continuar.

En primer lugar, lo más habitual es que el resultado del proceso no sea una respuesta en blanco o negro, y se requiera una decisión basada en un juicio clínico (situación que, a su vez, puede ser fuente de sesgo)[13]. Ello puede representar un inconveniente para algunos profesionales de la salud, quienes, en su deseo de tomar decisiones clínicas, buscan respuestas que les permitan saber si el artículo "está bien o mal", si el tratamiento allí descrito "sirve o no" y si los resultados "se aplican o no" a sus pacientes. Desafortunadamente los resultados de la evaluación de calidad de la investigación vienen en tonos de gris, y, por tanto, en muchas ocasiones a lo sumo se puede decir que los resultados pueden ser válidos, que su efecto, quizá, es importante, y, por tanto, podrían mejorar el cuidado en salud. A menudo, la evaluación de los estudios es muy útil si permite, al menos, asegurar que no hay un defecto metodológico preponderante o evidente, y que no hay razones para pensar que los resultados no son aplicables a otras poblaciones.

En segundo término, es necesario tomar en cuenta que no siempre los defectos metodológicos o los aspectos criticables hallados tienen un impacto real en los resultados del estudio; o sea, no siempre la existencia de una falla invalida las conclusiones. Como no hay ningún estudio perfecto, la intención no es descartar todo cuanto tenga alguna falla y aceptar solo lo que no tenga defectos. Todos los estudios publicados pueden tener defectos menores (incluso, moderados), pero su existencia no necesariamente invalida los resultados. El uso exagerado de los criterios de apreciación crítica puede llevar a que nunca se encuentre información de suficiente validez en la cual se puedan basar las decisiones clínicas.

EJEMPLO

El ya mencionado ECA sobre el hipolipemiante y su efecto en la prevención del infarto de miocardio puede haber tenido, como resultado del azar, una diferencia apreciable en la distribución por sexos entre los grupos. Mientras que en el grupo del medicamento el 65 % de los sujetos son hombres, el mismo porcentaje en el grupo de placebo es del 51 %. En este caso, como hay mayor riesgo cardiovascular en los hombres, antes de empezar el estudio hay una asimetría, pues el grupo que recibirá el medicamento tiene mayor riesgo de base. Sin embargo, la diferencia *per se* no anula el resultado. Como el estudio mostró que el hipolipemiante redujo de manera significativa el riesgo, se puede aceptar la dirección del resultado, por cuanto se observó que el medicamento produjo un mejor resultado que el placebo, incluso en un grupo que tenía mayor riesgo al iniciar el estudio; es decir, pese a la diferencia en los riesgos basales entre los grupos del estudio, puede aceptarse que el medicamento reduce de manera significativa el riesgo. Por otra parte, no ocurre lo mismo con la magnitud del resultado. Dados el mayor riesgo basal en el grupo expuesto al medicamento y la reducción observada en la frecuencia de infarto, puede inferirse que el efecto real debe de ser mayor que el observado. En otras palabras, se puede aceptar el resultado numérico, pero debe tomarse en cuenta que el verdadero efecto debe de ser aún mejor.

EVALUACIÓN DE LA VALIDEZ INTERNA

Las **tablas 35.1** y **35.2** presentan los principales aspectos metodológicos que deben ser evaluados por los investigadores de cada uno de los principales tipos de publicaciones[1,8-10]. Para ilustrar los conceptos, se describirá en mayor detalle la guía que evalúa un ECA que estudia el efecto de una intervención preventiva, o terapéutica. Algunos de estos aspectos son aplicables a los otros tipos de publicaciones.

Como lo muestra la **tabla 35.1**, el primer aspecto que se debe considerar para la pregunta de investigación del caso es si al iniciar el estudio los grupos experimental y de control tenían pronóstico similar. Esto es crítico, pues el resultado final de las enfermedades depende de muchas condiciones, además del tratamiento: la severidad de la enfermedad, la presencia de procesos patológicos asociados, la edad, el sexo y muchos más; o incluso, factores que pueden no haber sido identificados.

Para determinar si los grupos tenían pronóstico semejante al inicio del estudio, se sugiere buscar en el reporte información sobre cuatro aspectos básicos: 1) la asignación aleatoria, 2) la asignación ciega, 3) el análisis por intención de tratar y 4) la confirmación de una distribución equitativa de variables pronósticas medidas al ingreso. El lector interesado en profundizar en estos conceptos puede consultar el capítulo sobre experimentos controlados aleatorios.

Para lograr que los grupos sean similares al iniciar el estudio, el investigador debe haber asignado la intervención que cada participante recibe por medio del azar. Dicha aleatorización distribuye los factores tanto de buen como de mal pronóstico (sean conocidos o desconocidos) de manera similar en los dos grupos, de forma que su única diferencia es el tratamiento recibido. Y si los factores no quedan distribuidos de manera similar, las diferencias se deberán al azar, y no a un sesgo, lo que permite tener en cuenta esas diferencias en el análisis. Los estudios sin asignación al azar siempre tendrán limitaciones para separar el efecto del tratamiento que se investiga del atribuible a otras variables que pueden afectar la respuesta a los tratamientos.

EJEMPLO

El bajo peso al nacer es un fuerte predictor de mortalidad neonatal. La distribución balanceada de esta variable en los grupos de un estudio que compara dos intervenciones en neonatos eliminará el bajo peso como explicación para las diferencias que puedan encontrarse al finalizar el estudio. En otras palabras, la aleatorización reduce la probabilidad (como fuente de sesgo) de una distribución no balanceada de los sujetos en los grupos del estudio.

Además de la distribución aleatoria, el investigador debe garantizar la asignación ciega (actualmente se prefiere la denominación *enmascarada)*. Esto se logra cuando la decisión sobre incluir a un sujeto en el estudio se toma sin conocer el tratamiento que dicho individuo recibiría en caso de participar. Lo contrario puede introducir sesgos, al crear grupos con diferencias al comienzo del estudio.

EJEMPLO

Un investigador puede decidirse a invitar a todos los sujetos (con buen o mal pronóstico) cuando él bien sabe que recibirán el tratamiento control, y no invitar a los del más pobre pronóstico, si anticipa que el tratamiento recibido será el experimental. Tal aspecto ha demostrado influir sobre los resultados de los estudios: los experimentos sin asignación enmascarada tienden a sobreestimar el efecto de la terapia experimental al comparar con lo observado en estudios con adecuado enmascaramiento[14,15].

Los dos métodos más comúnmente utilizados para la asignación enmascarada son la aleatorización mediante llamada telefónicas a un centro coordinador y el empleo de sobres sellados, opacos y numerados secuencialmente, que contienen el tratamiento y solo deben abrirse una vez el sujeto ha sido aceptado en el estudio. Dicha asignación ciega no debe confundirse con la ***evaluación ciega de los desenlaces***, que se describe más adelante.

Luego de ser asignados, los sujetos deben ser analizados en su grupo original de asignación, así nunca hayan recibido la intervención intentada, o, incluso, si reciben la intervención alterna. Ello puede ocurrir porque el sujeto olvida usar el tratamiento, lo suspende debido a efectos adversos o por errores en la ejecución del estudio.

Cualquiera que sea la razón, para el análisis se debe mantener la asignación original empleando la estrategia conocida como ***análisis*** *por intención de tratar*. Hacer lo contrario puede introducir sesgo por dos razones básicas. En primer lugar, excluir a dichos individuos del estudio o cambiarlos de grupo equivaldría a omitir la asignación aleatoria. En segundo término, porque está ya claramente demostrado que el pronóstico de los sujetos que reciben los tratamientos es diferente del pronóstico de quienes lo suspenden, incluso cuando dicho tratamiento es un placebo[11].

El análisis por intención de tratar (lo contrario se llama ***análisis por protocolo***) acerca los experimentos a la vida real, pues se espera que en la vida real los individuos incumplan algunas veces las instrucciones para tomarse los medicamentos, que no se los tomen de manera completa, o, incluso, que no se los tomen.

Por último, los autores deben mencionar si los grupos del estudio eran similares antes de las intervenciones; o sea, mostrar que la asignación aleatoria logró, en efecto, uno de sus cometidos: producir grupos balanceados en los aspectos conocidos (los otros cometidos son lograr el balance en los aspectos no conocidos, evitar sesgos en la asignación y, por supuesto, producir variables aleatorias). Esta información suele presentarse en la primera tabla del artículo (tabla No. 1 del estudio), y les permite a los lectores establecer la dirección y la magnitud de las diferencias iniciales, que podrían confundir el efecto del tratamiento.

No obstante lo anterior, poco se logra comenzando el experimento con grupos que tienen el mismo pronóstico si a lo largo de la ejecución del estudio se presentan circunstancias que pueden introducir sesgo. Por ello, el lector debe establecer si el pronóstico se mantuvo comparable entre los grupos experimental y de control después de empezado el estudio[1,8-10]. Para asegurar ese aspecto, tanto los investigadores como los sujetos participantes y sus médicos tratantes han debido permanecer enmascarados respecto al tratamiento recibido. Y también, para asegurarse de que se mantiene la comparabilidad entre los grupos, los investigadores deben haber podido seguir con éxito a todos los individuos asignados (o, al menos, a la gran mayoría); y si hay pérdidas, no debe haber diferencias significativas en las pérdidas por grupo. A esa situación se le llama ***pérdida diferencial.***

Mantener a los investigadores, a los pacientes participantes y a los clínicos tratantes enmascarados respecto al tratamiento recibido reduce la probabilidad de sesgo en el momento de establecer si un individuo presentó el desenlace de interés. Hay evidencia empírica que muestra que si no se ha hecho el estudio enmascarado cuando debería serlo, los resultados del experimento pueden sobreestimar el efecto de la intervención en estudio[16]. Ello es particularmente crítico si la determinación del desenlace se basa en evaluaciones subjetivas por parte de un observador (por ejemplo, interpretar

una imagen diagnóstica, una prueba psiquiátrica o determinar la causa inmediata de muerte) o del paciente mismo (cambios en síntomas como el dolor, el prurito o las náuseas).

Si el observador conoce el tratamiento recibido y hay alguna expectativa por su parte respecto a la efectividad de las intervenciones comparadas, es posible que interprete de manera distinta los resultados de sujetos de un grupo en relación con el otro. Por otra parte, el efecto placebo (el cual se describe en detalle en el capítulo del ECA) puede hacer que un individuo reporte modificaciones en sus síntomas como consecuencia de las expectativas puestas en el tratamiento. La mejor manera de eliminar dichas fuentes de error es no informar a los observadores o a los pacientes el tratamiento que están recibiendo. Finalmente, si los clínicos tratantes conocen el tratamiento recibido, es posible que agreguen a los sujetos de uno de los grupos otras intervenciones que pueden modificar el desenlace, lo cual, a su vez, puede contribuir a diferencias en los resultados. Algo similar ocurre si el seguimiento es más común en un grupo que en el otro, pues es posible que se encuentren más eventos en el grupo seguido más de cerca. En ambos casos se presentará una forma de sesgo conocida como ***cointervención.***

El último aspecto que puede cambiar el pronóstico de los grupos durante el seguimiento es la pérdida de participantes durante el estudio[1,8-10]. En condiciones ideales, todos deberían ser seguidos para determinar su resultado definitivo. El problema virtual es que el investigador no puede establecer el resultado final para los individuos perdidos, y es muy factible que la pérdida se asocie a la presencia de uno o más desenlaces de interés (como ya se había comentado, puede ser que la pérdida se deba a que los pacientes murieron [***variable de interés***], o se recuperaron definitivamente [también desenlace de interés] o presentaron efectos adversos severos [***desenlace*** *de interés en seguridad*]).

Otra analogía clínica ayuda a explicar este punto. A veces, los clínicos pierden de vista a sus pacientes luego de recomendarles un tratamiento. En tales circunstancias, lo más probable es que la condición del sujeto haya evolucionado (mejoró o empeoró), de forma que el mismo sujeto considera innecesario regresar para controles. En otras palabras, los pacientes desaparecen porque presentan desenlaces desfavorables (incluso, la muerte), o bien, porque mejoran. Desafortunadamente, no hay un porcentaje "seguro" de sujetos perdidos.

La forma de establecer si las pérdidas pueden afectar los resultados es adelantar un segundo análisis contando a los sujetos perdidos, los del grupo que presenta los mejores resultados, como desenlaces ***desfavorables***; y a los perdidos en el grupo con los peores resultados, como *favorables*. Si tal estrategia produce resultados similares a los reportados sin considerar las pérdidas, se puede concluir que estas, probablemente, no introdujeron ningún sesgo. Por otra parte, los hallazgos pierden credibilidad si los resultados obtenidos en los dos tipos de análisis son diferentes.

EJEMPLO

Un investigador evalúa un antibiótico nuevo, la todomicina, y la compara con un antibiótico convencional en el tratamiento de infección urinaria no complicada en 200 pacientes. Al terminar el estudio, 95 de los 100 sujetos asignados al antibiótico convencional pudieron ser seguidos con éxito, en tanto que solo 71 de los 100 sujetos asignados fueron seguidos con éxito en el grupo de todomicina. En este caso, la diferencia entre los dos grupos en el porcentaje de pérdidas sugiere la existencia de factores no identificados que hacen que quienes reciben el tratamiento experimental tengan menos adherencia al tratamiento, o mueran más, o mejoren más (y por eso no vuelvan al estudio), o tengan más reacciones adversas que los impulsen a no seguir participando en el estudio.

EVALUACIÓN DE LA IMPORTANCIA DE LOS RESULTADOS

Luego de verificar que los métodos empleados por los investigadores fueron apropiados, es necesario evaluar el impacto potencial de los hallazgos del estudio. Dicho impacto potencial depende del tamaño y de la precisión del efecto reportado en el artículo. Esos conceptos se explicarán en el caso de un estudio que compara dos intervenciones, aunque son aplicables a los restantes tipos de estudios.

El objetivo final de la atención en salud es mejorar el bienestar de los individuos reduciendo la probabilidad de desarrollar enfermedades a través de intervenciones preventivas, o reduciendo el riesgo de desenlaces indeseables (la muerte, la recaída o el desarrollo de com-

plicaciones o discapacidad) si la enfermedad ya está presente, a través de diversas formas de tratamiento. De esa manera, para tener un impacto positivo en la salud de los individuos, un estudio debe demostrar que el riesgo de la enfermedad o de resultados indeseables es menor en el grupo que recibe la nueva intervención cuando se lo compara con el riesgo en el grupo que recibe el tratamiento convencional.

El **capítulo 9** "Medidas de frecuencia, de asociación y de impacto" detalla las medidas de resumen para comparar los riesgos de los grupos del estudio. Quizás, la más común es el ***RR***, que es la razón entre la frecuencia del desenlace de interés en el grupo expuesto al nuevo tratamiento, en relación con la del grupo control[1,8-11]. Una razón (**RR**) de 1,0 indica que los riesgos son iguales en los dos grupos. Los resultados < 1 sugieren que el riesgo es menor en quienes reciben la intervención; por ejemplo, un **RR** de 0,5 indica que la frecuencia de eventos indeseables en el grupo experimental fue la mitad de la del grupo control (esto equivale a una reducción del 50 % en el riesgo). Por el contrario, un **RR** > 1,0 indicaría que el riesgo es mayor entre los sujetos expuestos a la nueva intervención; un valor de 1,7, por ejemplo, significa que el riesgo es 1,7 veces (o el 70 %) mayor entre los expuestos, cuando se lo compara con el riesgo en los sujetos no expuestos. De esa forma, cuanto más pequeño sea el **RR**, mayor será el tamaño del efecto positivo (o cuanto mayor sea el **RR**, mayor será el tamaño del efecto negativo) de la exposición de interés sobre el desenlace.

El tamaño del efecto, no obstante lo anterior, no depende exclusivamente del resultado numérico del **RR**. Otros factores que deben tenerse en cuenta incluyen la importancia clínica del desenlace y su frecuencia en los sujetos que no reciben tratamiento. Estos dos puntos serán discutidos con más detalle más adelante en este mismo capítulo.

El segundo aspecto que debe considerarse durante la evaluación del impacto potencial de la intervención es la precisión del resultado. Para comprender este concepto, es necesario recordar el efecto del azar, o *error aleatorio*, en la medición repetida del peso de un niño, ya descrita. De manera similar, el **RR** obtenido al final del estudio podría estar afectado por variación aleatoria: el resultado del estudio podría ser mayor o menor que el reportado (y, por ende, mayor o menor que el verdadero valor, que se trata de estimar con el estudio) si se repite el experimento muchas veces (algo que, en realidad, no se hace). En otras palabras, es poco probable que el resultado estimado en el estudio represente con exactitud el verdadero tamaño del efecto.

Una medida de la variabilidad que puede esperarse por efecto del azar es el *intervalo de confianza* (**IC**); usualmente, del 95 % (**IC95 %**) alrededor del valor encontrado (el lector interesado puede consultar el **capítulo 11** "La propuesta de investigación"). Este intervalo se interpreta de la siguiente forma: en un estudio sin sesgos hay el 95 % de probabilidad de que el verdadero tamaño del efecto esté en algún lugar entre dichos límites[1,8,11].

EJEMPLO

En un experimento clínico que encontró un **RR** para el desenlace de interés (muerte) de 0,5; los **IC95 %** alrededor del riesgo relativo fueron 0,35-0,8. Esto significa que se tiene un indicador general de 0,5 como **RR** (reducción de la mortalidad en el 50 %), y se tiene el 95 % de certeza de que el verdadero tamaño del efecto está en algún lugar entre esos límites: 0,35 (reducción del 65 %) y 0,80 (reducción del 20 %).

Cuanto más estrecho sea el **IC**95 %, mayor será la certeza del investigador y del lector respecto al verdadero tamaño del efecto. Dicha precisión estadística depende, principalmente, del poder estadístico del estudio, que, a su vez, depende, sobre todo, del tamaño de la muestra. Esto confirma algo que es correcto intuitivamente para la mayor parte de los clínicos: cuanto mayor sea la muestra de un estudio, mayor será, en general, la credibilidad de sus resultados.

Además de medir el error aleatorio, los **IC** contienen información equivalente a la de las pruebas estadísticas para hipótesis. Si el valor del **RR** de 1,0 (el cual señala que no hay diferencias entre los riesgos) está incluido dentro de los límites del **IC95 %** (por ejemplo, un **RR** de 0,5 con **IC95 %** entre 0,2 y 1,1), es muy improbable que los resultados alcancen una diferencia estadísticamente significativa a nivel de $p < 0,05$. Si el **IC** no incluye 1,0, los resultados son estadísticamente significativos (por ejemplo, en el intervalo de 0,35-0,8).

Los **IC** son una herramienta muy poderosa para interpretar los resultados de la investigación, pues permiten evaluar el rango de posibles resultados compatibles con los hallazgos del

estudio. Eso puede ayudar al lector a definir si un tamaño de efecto considerado clínicamente importante es consistente con los datos, o bien, si puede ser descartado por estos.

EJEMPLO

Un RR de 0,96 sugiere una reducción de tan solo el 4 % en el riesgo del desenlace como consecuencia de la intervención. Si el IC95 % fuera de 0,9-1,1, sería casi seguro que el estudio ha descartado un tamaño de efecto clínicamente importante. La interpretación sería del todo diferente si los límites fueran de 0,3-2,1. En este caso, los resultados son compatibles tanto con una reducción del 70 % como con un aumento del 110 % en el riesgo del desenlace en los sujetos expuestos a la intervención. Debido a un bajo poder, el mencionado estudio no ha descartado una diferencia clínicamente importante entre los tratamientos, y, por tanto, no puede llamarse un estudio "negativo". Un mejor término para tal situación sería *inconclusivo*.

EVALUACIÓN DE LA GENERALIZABILIDAD O DE LA VALIDEZ EXTERNA

El último paso en la evaluación crítica de un artículo es establecer si los resultados le ayudan al lector a brindar un mejor cuidado a sus pacientes. Este punto es de particular importancia para los clínicos que trabajan en los países menos desarrollados, pues la mayor parte de los estudios publicados han sido conducidos en países industrializados. Uno de los números de la serie "Users' Guides to the Medical Literature" se dedica a este aspecto[1,16]. La combinación de aspectos biológicos, socioeconómicos y epidemiológicos determina si los resultados de un estudio son aplicables a un individuo.

En relación con los aspectos biológicos, puede haber diferencias fisiopatológicas en la enfermedad, entre diversos lugares del mundo, que pueden modificar la respuesta al tratamiento. Este es el caso de algunas enfermedades infecciosas que pueden mostrar diferencias en la susceptibilidad del agente causal a los antibióticos.

Por otra parte, puede haber diferencias en las poblaciones en aspectos como el metabolismo de los fármacos, la respuesta inmune o las condiciones ambientales que afecten la efectividad y la seguridad de las intervenciones. Los aspectos sociales y económicos incluyen variaciones en la adherencia, tanto de los pacientes como de los profesionales de la salud, que pueden llevar a modificaciones en la forma como se prescriben o se usan los tratamientos. Tal variabilidad a menudo surge como consecuencia de limitaciones económicas, de dificultades en el acceso a recursos técnicos, de diferencias en las habilidades del personal de salud y de la idiosincrasia del comportamiento (por ejemplo, en Colombia es común la creencia de que los broncodilatadores inhalados producen dependencia, lo cual lleva a que a menudo los pacientes suspendan su uso).

Finalmente, la alta frecuencia de otras condiciones mórbidas en un área geográfica particular puede afectar la efectividad de la intervención, al hacer más difícil el diagnóstico antes de iniciar la terapia o por aumento en el riesgo de desenlaces adversos. Un buen ejemplo es la desnutrición infantil, la cual hace más difícil la identificación de casos con neumonía, y, a su vez, aumenta el riesgo de una pobre respuesta terapéutica.

Es claro, a partir de lo dicho en los párrafos precedentes, que establecer si los resultados de un estudio en particular pueden ser aplicados a otros sujetos es materia digna de un juicio clínico cuidadoso. Este es un aspecto de la apreciación crítica donde no hay una respuesta "correcta". Como regla general, se sugiere que el lector se pregunte si el paciente a quien está viendo podría haber ingresado al estudio en caso de haberse hallado en el centro donde se adelantó la investigación[1,8,16]. Si el paciente cumple con los criterios de inclusión y no tiene ninguno de los de exclusión, es razonable concluir que, muy probablemente, los resultados le sean aplicables. En caso contrario, el lector debe pensar que hay razones de peso por las cuales los resultados no deberían aplicarse a dicho individuo.

Es necesario considerar otros factores que determinan si los resultados de un estudio serán de utilidad para el lector con el fin de brindar una adecuada atención a sus pacientes. Dos aspectos, ya mencionados, son la importancia clínica del desenlace y su frecuencia cuando se emplea el tratamiento convencional.

Para modificar la práctica, los lectores deben buscar evidencia de cambios de magnitud clínicamente significativa en desenlaces que, a su vez, sean clínicamente importantes. No es raro

que los investigadores empleen ***desenlaces sustitutos***, que son más fáciles de medir o reducen el tamaño de la muestra necesario para el estudio. En enfermedades como el asma bronquial y o la bronquiolitis, se emplean, por ejemplo, puntajes obtenidos a partir de síntomas y signos. Sin embargo, demostrar que hay diferencias en esos parámetros entre los grupos del estudio no necesariamente significa que los pacientes van a experimentar una reducción en el riesgo de otros desenlaces de mayor importancia clínica, como la muerte, la recaída, la necesidad de hospitalización o la frecuencia de complicaciones.

Por otra parte, el lector debe también buscar información sobre la frecuencia y la severidad de los efectos secundarios asociados a la nueva intervención, como parte de la ***evaluación de*** *validez externa*, o capacidad para generalizar los resultados. La pregunta final respecto a si adoptar una nueva intervención o una nueva prueba diagnóstica conlleva establecer si los beneficios probables que se derivan de su empleo son superiores a los riesgos y a los costos asociados. Un aspecto importante en tal sentido es la frecuencia del desenlace que se quiere evitar, en sujetos que no reciben la intervención nueva, sino la convencional, lo que se conoce como el ***riesgo basal***.

EJEMPLO

Ya se tienen vacunas efectivas contra el rotavirus, la causa más común de diarrea en la población infantil. Si se asume que dicha vacuna estará disponible en cualquier lugar del mundo, es intuitivamente obvio que su impacto será mayor en los lugares menos desarrollados, debido no solo a la mayor carga de la enfermedad, sino al mayor riesgo de complicaciones y de muerte en los niños que desarrollan diarrea en las regiones más pobres, cuando se compara con el riesgo observado en países con mayor desarrollo.

Una medida útil para evaluar el impacto potencial de una intervención consiste en estimar el número de sujetos que deben recibir la intervención para evitar un desenlace indeseable. A este número se le conoce como ***número necesario para tratar***, o NNT (que también se describe en detalle en el **capítulo 9** "Medidas de frecuencia, de asociación y de impacto"). El NNT es el inverso de la diferencia entre el riesgo del evento observado en el grupo control y el grupo experimental; o, lo que es igual, el recíproco de la diferencia absoluta en el riesgo[1,8]. Cuanto más pequeño sea el NNT, mayor será el impacto de una intervención, pues se requiere que menos pacientes la reciban para prevenir un caso o un desenlace adverso.

CONSECUENCIAS (BUENAS Y MALAS) DE LA LECTURA CRÍTICA DE LA LITERATURA MÉDICA

Las consideraciones finales de este capítulo se relacionan con las ventajas y las desventajas de la lectura crítica de la literatura médica. Las consecuencias positivas incluyen un uso más eficiente del tiempo disponible para la lectura, bien sea dirigida a resolver problemas clínicos o a mantenerse actualizado.

Los pasos sugeridos en las secciones precedentes le permitirán al lector ocupado determinar con prontitud si el artículo que tiene en sus manos le brindará información válida y aplicable. Si ese no es el caso, el lector puede abandonar la lectura del artículo, para pasar a la siguiente publicación. Por otra parte, hay evidencia, aunque de calidad limitada y variable, indicativa de que enseñar apreciación crítica a los estudiantes de medicina durante los últimos años de su formación, así como a los residentes y a los médicos ya graduados, puede mejorar sus habilidades para el diagnóstico y el tratamiento; al menos, en casos escritos[17-19]. Adicionalmente, el lector tendrá mejores elementos para entender los principios de una buena investigación en salud. Por último, los clínicos dispondrán de evidencia más sólida para el proceso de toma de decisiones.

Pese a lo anterior, hay algunas desventajas asociadas a la lectura crítica de la literatura. La principal es que la actividad consume tiempo, no solo el necesario para adquirir los conocimientos básicos, sino también, para desarrollar las habilidades necesarias. Como es usual en toda actividad que requiere el desarrollo de destrezas, estas se adquieren y se mantienen solo a través de la práctica. La lectura de textos como este capítulo es únicamente el primer paso, en el mismo sentido en que leer la técnica para hacer una endoscopia digestiva es un punto de partida apropiado, pero insuficiente, para llevar a cabo el procedimiento con éxito. La práctica y el tiempo harán el proceso más simple y agradable.

Adicionalmente, el lector puede sentirse frustrado al hallar deficiencias en la mayoría de los artículos que lee, lo cual se puede explicar

por dos razones: 1) o bien, le falta madurez científica y metodológica a la revista que publicó el estudio (lo cual es aún muy habitual; sobre todo, en países en desarrollo y en revistas secundarias, sin evaluación por pares), 2) o bien es el resultado de una aplicación muy rígida de los parámetros de lectura crítica.

Por último, en los ambientes donde el concepto de los *mayores* (profesores de mayor prestigio y edad) es el elemento de más importancia en la toma de decisiones, el ejercicio de apreciación crítica puede representar una amenaza a su autoridad y al sistema.

Puede concluirse que hay un método sistemático y con fundamento para el análisis crítico de la literatura médica publicada, y el cual se basa en los principios metodológicos que garantizan resultados confiables y aplicables. El ejercicio razonable de esta técnica ha mostrado tener un impacto positivo en la atención médica en salud y en la eficiencia en el uso del tiempo del médico clínico.

Referencias

1. Guyatt G, Rennie D, Meade M, et al, editors. Users´ Guides to the medical literature: essentials of evidence-based clinical practice. 2nd ed. Nueva York, NY: McGraw-Hill; 2008.
2. Haynes RB, McKibbon KA, Fitzgerald D, et al. How to keep up with the medical literature: I. Why to try to keep up and how to get started. Ann Intern Med. 1986;105:149-53.
3. Moher D, Dulberg CS, Wells GA. Statistical power, sample size and their reporting in randomized controlled trials. JAMA. 1994;272:122-4.
4. Harhay MO, Wagner J, Ratcliffe SJ, et al. Outcomes and statistical power in adult critical care randomized trials. Am J Respir Crit Care Med. 2014;189:1469-78.
5. Schulz KF, Chalmers I, Grimes DA, et al. Assessing the quality of randomization from reports of controlled trials in obstetrics and gynecology journals. JAMA.1994;272:125-8.
6. Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomized controlled trials: survey of published parallel group trials in obstetrics and gynaecology. BMJ. 1996;312:742-4.
7. loannidis JPA. Why most published research findings are false. PLoS Medicine. 2005;2:e124.
8. Straus SE, Glasziou P, Richardson WS, et al. Evidence-Based medicine. How to practice and teach EBM.4th ed. Edimburgo: Elsevier Churchill Livingstone; 2011.
9. Riegelman RK. Studying a study and testing a test. Reading evidence-based health research. 6th ed. Philadelphia: Wolters Kluwer - Lippincott Williams & Wilkins; 2012.
10. Greenhalgh T. How to read a paper. The basics of evidence based medicine. 5th ed. London: Wiley Blackwell - BMJ Books; 2014.
11. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: The Essentials. 5th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2014.
12. Fitzgerald FT. From galen to xerox: the authoritarian reference in medicine. Ann Intern Med. 1982;96:245-6.
13. Kaptchuk TJ. Effect of interpretative bias on research evidence. BMJ. 2003;326:1453-5.
14. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273:408-12.
15. Moher D, Pham B, Jones A, et al.Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609-13.
16. Dans AL, Dans LF, Guyatt GH, et al. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. JAMA. 1998;279:545-9.
17. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. CMAJ. 1998;158:177-81.
18. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. Acad Med. 1999;74:686-94.
19. Horsley T, Hyde C, Santesso N, et al. Teaching critical appraisal skills in healthcare settings. Cochrane Database Syst Rev. 2011:CD001270.

CAPÍTULO

36

Principios éticos en investigación clínica y epidemiológica

Álvaro J. Ruiz Morales • Fernando Suárez Obando

INTRODUCCIÓN

La investigación y la ética deben estar siempre ligadas, de manera estrecha e inseparable, puesto que los principios éticos son aspectos fundamentales de la investigación que involucrea a seres vivos. Esta relación entre ciencia y reflexión moral es más relevante en la investigación en seres humanos, en todos los casos; pero se convierte en un tema especialmente sensible cuando los sujetos involucrados están en condiciones de vulnerabilidad.

Cuando se evalúa, bien por parte del investigador, o bien, por parte de un comité de investigación y ética, un protocolo de investigación con seres humanos en el área clínica, deben tenerse en mente los tres aspectos básicos de la investigación: 1) el área temática, 2) la metodología y 3) la ética.

Y debe evaluarse teniendo en cuenta que la ética comprende transversalmente todo el proceso investigativo, y que sus principios se deben reflejar en el abordaje metodológico que pretende resolver una pregunta científica sólida y válida, que con frecuencia debe ser resuelta a través de la vigilancia o de intervenciones experimentales en personas en estado de vulnerabilidad, como en el caso de personas en riesgo o que padecen una enfermedad.

Es claro que la concepción de la investigación debe ir más allá de los principios básicos éticos reglamentarios y debe considerar el entorno particular en cada situación y sus posibles implicaciones. A este enfoque global y más amplio se le conoce como ***bioética.***

Infortunadamente, no siempre se ha considerado necesario hacer una evaluación bioética de la investigación. De hecho, las reglamentaciones que establecen tener en cuenta los derechos de los sujetos de investigación y los deberes de los investigadores y de las instituciones con dichos sujetos son relativamente recientes; en gran medida, como resultado de los juicios a los crímenes cometidos por médicos nazis con prisioneros durante la Segunda Guerra Mundial.

El Tribunal Internacional Militar, que juzgó a 20 médicos acusados de asesinato y tortura en los campos de concentración, además de adelantar el juicio, consideró aspectos mucho más amplios respecto a la forma correcta de conducir investigación en seres humanos. Esto llevó a que en agosto de 1947 se promulgara una serie de principios sobre la investigación en humanos, conocidos como el Código de Núremberg[1]. Más adelante, y con base en ese código, la Aso-

ciación Médica Mundial estableció en 1954 los principios éticos para la investigación en seres humanos, que evolucionaron hasta dar lugar, en medio de la XVIII Asamblea Médica Mundial, en 1964, a la Declaración de Helsinki, ratificada y actualizada en varias oportunidades; la última, en 2013, en Fortaleza, Brasil[2].

En 1966, la Asamblea General de las Naciones Unidas adoptó el Pacto Internacional sobre Derechos Civiles y Políticos, que estableció varios principios importantes como norma internacional; entre ellos, el artículo 7: "Ninguna persona será sujeta a tortura o a castigo o tratamiento cruel, inhumano o degradante. En particular, ninguna persona será sujeta sin su libre consentimiento a experimentación médica o científica"[3].

A partir de los años sesenta del siglo XX, los Institutos Nacionales de Salud (NIH, por las siglas en inglés de National Institutes of Health), la Administración de Drogas y Alimentos (FDA, por las siglas en inglés de Food and Drugs Administration) y otras entidades de Estados Unidos, así como el Consejo para Organizaciones Internacionales de Ciencias Médicas (CIOMS, por las siglas en inglés de Council for International Organizations of Medical Sciences) y la Organización Mundial de la Salud (OMS), han adoptado y ampliado esos elementos fundamentales, con lo cual se dio lugar a un marco amplio de principios éticos, que actualmente orientan la investigación que involucra a sujetos humanos. En particular, en 1978 se publicó el Informe Belmont, creado por la Comisión Nacional para la Protección al Sujeto Humano en Investigación Biomédica y Comportamental de los Estados Unidos (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research)[4], que promulga claramente la observación de tres principios fundamentales en investigación: 1) el respeto por las personas, 2) la beneficencia y 3) la justicia; ello inaugura un marco principialista que ha dominado la ética de la investigación a partir de su publicación.

Posteriormente, en 1993, y en razón de circunstancias especiales en los países en vías de desarrollo respecto a la aplicabilidad del Código de Núremberg y de la Declaración de Helsinki, del principialismo y de la investigación clínica desarrollada, principalmente, por las multinacionales farmacéuticas, la organización CIOMS y la OMS expidieron las *Guías internacionales para la investigación biomédica que involucre a seres humanos*, conocidas también como las Pautas CIOMS[5], pautas que han sido complementadas y ampliadas por diversos documentos sobre ética de la investigación publicados por CIOMS y por otras organizaciones, como la OMS y la Organización Panamericana de la Salud (OPS).

El planteamiento de CIOMS contrasta con el documento de las buenas prácticas clínicas de 1996, elaborado por la Conferencia Internacional de Armonización (GCP-ICH, por las siglas en inglés de Good Clinical Practice-International Conference on Harmonisation), que buscó estandarizar la práctica investigativa para los procedimientos de investigación necesarios para el estudio de medicamentos de la industria farmacéutica[6], planteamiento que, junto con el principalísimo, ha dominado el enfoque pragmático de la ética de la investigación durante las últimas décadas.

PRINCIPIOS ÉTICOS DE LA INVESTIGACIÓN

El enfoque basado en principios comprende al menos tres principios éticos básicos en relación con la investigación en sujetos humanos: 1) el respeto por las personas, 2) la beneficencia y 3) la justicia; sin embargo, se han desarrollado otros marcos éticos dentro de los cuales es posible tener una aproximación a la ética de la investigación, que, además, de comprender el principialismo clásico, también abarca aspectos del contexto investigativo y su impacto social.

A continuación se exponen los tres principios básicos, y, posteriormente, un marco ampliado de principios aplicado a la investigación clínica.

Respeto por las personas

El respeto incluye dos principios relacionados: 1) la autonomía de cualquier individuo y 2) la protección de personas con autonomía disminuida.

Una persona autónoma es aquella "capaz de deliberar sobre sus objetivos personales y de actuar bajo la dirección de tal deliberación". El proceso del consentimiento informado proviene directamente de este principio.

El respeto por las personas requiere, además, que quienes sean vulnerables o dependientes (se verán casos específicos más adelante) reciban protección especial contra riesgos o abuso, que tengan un marco concreto de propuestas que los protejan en su participación en estudios clínicos y que no se vean privados de su derecho a participar en proyectos de investigación.

Beneficencia

Es la obligación ética de llevar al máximo los posibles beneficios y reducir a su mínima expresión los riesgos.

Este concepto se basa en el axioma hipocrático conocido como ***primum non nocere***, o ***ante todo no hacer daño***, y exige que los riesgos para los sujetos de investigación sean razonables en relación con los beneficios esperados, que el diseño de la investigación sea apropiado y que los investigadores sean competentes para conducir la investigación y para asegurar el bienestar de los participantes.

Idealmente, el mencionado principio debe desprenderse del beneficio inmediato potencial para el sujeto que participa en un estudio, además del beneficio que se espere para la comunidad.

El principio de beneficencia debe ser complementado con el concepto de ***no maleficencia***, más apropiado en cuanto a su sentido. Si bien se debe procurar siempre buscar el bien y llevar al máximo el beneficio, en ocasiones es necesario exponer al sujeto (y más tarde, al paciente) a algunos riesgos inevitables. Por esto, la descripción del principio como no maleficencia resulta más ajustada, ya que se debe buscar ofrecer beneficio, pero también es importante que no se busque hacer daño, y que si resultan inevitables ciertos riesgos, estos se acepten como parte de un balance adecuado con los beneficios.

Justicia

Requiere que los casos considerados similares sean tratados de forma parecida; es decir, que la distribución de los riesgos y de los beneficios de la investigación —especialmente, las cargas o sus riesgos— no se concentre en grupos con ciertas desventajas

(Los más pobres o minorías étnicas o raciales, por ejemplo) cuando las ventajas potenciales de la investigación benefician a todos. En otras palabras, los estudios deben ser diseñados para obtener conocimientos que beneficien a la clase de personas a quienes representan los sujetos participantes.

Estos tres principios determinan los restantes aspectos éticos de la investigación en seres humanos, que se traducen en diversas guías preparadas por el CIOMS y la OMS. Estos documentos consideran detalladamente los principios y las recomendaciones sobre aspectos como el consentimiento informado y sus peculiaridades en situaciones especiales (investigación con menores de edad, con sujetos que padecen trastornos mentales o del comportamiento, con prisioneros, con comunidades menos desarrolladas), los criterios éticos para seleccionar los sujetos de investigación, la confidencialidad de la información, el derecho a compensación para el participante en caso de lesión, el proceso de revisión por los comités de revisión ética y las obligaciones de los patrocinadores y de los países participantes, en caso de investigación internacional. Algunos incluyen anexos con la Declaración de Helsinki y las fases de los experimentos clínicos de vacunas y medicamentos.

Sin embargo, la reflexión moral acerca de la investigación clínica no se limita a los principios, sino que considera sus alcances a largo plazo, el contexto en que se lleva a cabo el desarrollo científico, los beneficios y los riesgos y una relación de doble vía entre los investigadores y los sujetos.

Por esta razón, los principios pueden plantearse dentro de consideraciones de mayor alcance que construyan un marco ético para la investigación científica[7].

MARCO ÉTICO PARA LA INVESTIGACIÓN

Un planteamiento más amplio de la ética de la investigación, que no implique solamente el supuesto cumplimiento o el apego a los principios generales y que no se limite al proceso de consentimiento informado, implica comprender el proceso investigativo en la complejidad de la relación entre sujeto e investigador, proceso que, a su vez, debe

estar plenamente justificado; especialmente, en relación con los riesgos y los beneficios derivados de la intermediación científica, y que se afirme en que la justificación de la pregunta científica no es suficiente para ejecutar un protocolo a costa de la dignidad de los sujetos.

Se plantea, entonces, que deben considerarse otros requerimientos que enmarquen una reflexión sobre el quehacer científico y superen la noción de requerimiento obligatorio que existe alrededor de la firma del consentimiento. Este marco ha sido desarrollo, principalmente, por Ezekiel J. Emanuel, entre otros[7,8] y responde a la necesidad ir más allá del principialismo y de tener como objetivo principal evitar la explotación de los sujetos involucrados en investigación. A continuación se resumen sus principales características.

Asociación colaborativa

El marco ético propone inicialmente que la investigación debe partir de una colaboración entre los investigadores, los patrocinadores, los responsables de las políticas de salud y la comunidad. Esta asociación debe involucrar a los participantes en acciones que determinen la importancia del problema de salud que se va a investigar; que evalúen el valor de la investigación a la luz del problema de salud identificado; que ejerzan liderazgos conjuntos en la planificación, la realización y la supervisión de la investigación, y que busquen y faciliten la integración de la investigación y de sus resultados al sistema sanitario.

Esta asociación debe propender por que las actividades de investigación respeten los valores de la comunidad, de la cultura, de las tradiciones y de sus prácticas sociales[9].

Al mismo tiempo, el trabajo colaborativo debe ser capaz de organizar la actividad de los investigadores en condiciones de igualdad frente a los responsables de las políticas de salud y frente a la comunidad, de tal modo que ni los científicos ni los patrocinadores tengan posiciones dominantes.

Por último, esta asociación colaborativa debe propender por que los participantes reclutados y las comunidades reciban los beneficios generados por los resultados de la investigación, así como por compartir los beneficios financieros de otro tipo que resulten de la investigación llevada a cabo.

Valor social

El valor social de la investigación debe especificar claramente los beneficiarios de la investigación, tanto a corto como a largo plazo, y una vez identificados dichos beneficiarios (por ejemplo, los pacientes, las comunidades, el país donde se reclutará a los sujetos), establecer cuáles serán el valor y la importancia de la investigación para cada grupo.

Se debe propender por la creación de mecanismos que potencien el beneficio de la investigación y diseminen los resultados a grupos clave, como pacientes, médicos, grupos de apoyo y legisladores de políticas en salud.

El desarrollo científico debe propender por fortalecer el sistema de salud estableciendo que un estudio de investigación debe ser factible dentro del contexto social, político y cultural, y apuntar hacia mejoras sostenibles en la infraestructura local del sistema de salud[10]. Esto significa que, idealmente, también los resultados del estudio deben poder ser aplicados a la comunidad en la que se desarrollaron.

Validez científica

La investigación debe diseñarse de tal forma que provea datos válidos y confiables, lo cual implica que el planteamiento metodológico y estadístico debe ser coherente con los objetivos planteados, los cuales, a su vez, deben ser claros y justificables.

Por otra parte, el diseño debe propender por que los resultados sean interpretables y útiles en el contexto clínico del problema que se está investigando.

Los objetivos del estudio deben tener en cuenta que para su cumplimiento no se debe negar la atención a un sujeto que aun hallándose fuera del estudio tendría derecho a ella, y se debe valorar la conveniencia de que reciba un manejo que de otra forma no sería posible recibir en el contexto sanitario del sujeto.

Por otra parte, es necesario, dentro del planteamiento del diseño investigativo, sopesar la conveniencia del uso de placebo, el cual solo se debe utilizar si se cumplen dos condiciones básicas: 1) que metodológicamente haya

razones suficientes para utilizarlo, y 2) que utilizarlo signifique una posibilidad ***mínima*** de hacer daño.

El diseño y la metodología científicos planteados en un protocolo, además de propender por la validez científica, deben tener en cuenta el bienestar y el respeto por los derechos de los sujetos y armonizar la validez científica, con el apego a los principios éticos, incluyendo la viabilidad del estudio en el contexto social, cultural y ambiental de lo sujetos a quienes se propone reclutar[11,12].

Selección justa de los participantes

La selección justa de los participantes surge de los criterios de elegibilidad de los sujetos y de la población en estudio, aspectos que, a su vez, se definen de acuerdo con los objetivos de la investigación.

Si la elegibilidad parte de los objetivos científicos, se evita que la inclusión de sujetos se deba a consideraciones externas al propósito investigativo, de tal modo que para cumplir con una selección justa, la población de estudio debe ser seleccionada con el fin de asegurar el propósito y la validez científicos; además, debe minimizar los riesgos a los sujetos y tomar en cuenta que la selección incremente el valor social de la investigación y defina posibles beneficios para los participantes.

Balance riesgo-beneficio favorable



La investigación clínica debe ofrecer a los sujetos de investigación una relación riesgo-beneficio favorable para los participantes.

Esta afirmación se hace extensiva al contexto del individuo participante, el cual define, en parte, lo que es benéfico para el sujeto en relación con los riesgos.

Este contexto incluye desde el ambiente y las condiciones sociales de la persona hasta sus determinantes biológicos y genéticos, aspectos que deben ser considerados por los investigadores en el momento de establecer el balance entre el riesgo de ser intervenido experimentalmente y el beneficio derivado de participar en la investigación.

Evaluación independiente



La revisión independiente del protocolo se hace necesaria para minimizar las preocupaciones alrededor de los conflictos de interés entre investigadores y patrocinadores, y mantener la transparencia, durante todo el proceso, del desarrollo del estudio.

Esta revisión abarca un proceso amplio de seguimiento, tanto de la ejecución de los protocolos como de verificación de la publicación de resultados luego de los análisis de resultados.

Este aspecto es esencial para orientar el funcionamiento del comité de ética, el cual, además de velar por el respeto al sujeto, debe identificar aspectos que desdibujen los propósitos científicos de la investigación.

Respeto por los participantes/ comunidades

El respeto por los participantes es un aspecto fundamental en el marco ético general, porque incluye, además de la observación y el respeto estricto de los derechos de los sujetos, la selección justa de los participantes y el balance riesgo-beneficio favorable a los participantes más allá del consentimiento informado, teniendo en cuenta el contexto comunitario del individuo. Es decir, para que el marco esté completo y se cumpla el objetivo de proteger al sujeto de investigación, el respeto por el participante es la culminación de un proceso que inicia con una adecuada fundamentación científica, la valoración de los riesgos, la selección justa de las personas participantes, la valoración de conflictos de interés y el contexto biológico y cultural de los sujetos reclutados.

Este proceso culmina en el consentimiento informado, el cual cierra el marco que engloba los principios éticos de la investigación científica.

Consentimiento informado



El consentimiento informado debe entenderse como un proceso, en el que se ratifica el deseo de una persona de participar; es la aceptación dada por un individuo competente, quien ha recibido del investigador

toda la información necesaria sobre el estudio, que ha comprendido adecuadamente la información y que después de considerarla, ha decidido participar sin haber sido sometido a coacción, a influencias inadecuadas ni a intimidación.

Por sí mismo, el consentimiento informado es una salvaguarda insuficiente para el sujeto participante, por lo que siempre debe complementarse con una revisión independiente sobre los propósitos de la investigación en la que se resalten los aspectos bioéticos.

Por otra parte, hay consenso en el mundo sobre los puntos mínimos que deben incluirse de forma explícita en el consentimiento **(tabla 36.1)**[13,14]. No sobra aclarar que el consentimiento informado debe considerarse un proceso que se inicia cuando se establece el primer contacto con un sujeto potencialmente elegible para un estudio, y que continúa a lo largo de este; por eso, el investigador tiene la obligación de contestar cualquier pregunta que un sujeto pueda hacerle durante el estudio, mientras que el participante conserva el derecho a retirar su consentimiento cuando quiera hacerlo, así haya aceptado previamente, y sin que este retiro le signifique ninguna desventaja ni penalidad.

Aunque en principio es obligatorio obtener consentimiento informado en todas las investigaciones, hay circunstancias especiales en las cuales no es práctico obtenerlo de cada sujeto o por escrito. Como ejemplos, en algunas poblaciones indígenas la autorización del jefe de la tribu es suficiente en esa cultura y no es necesario ni práctico pedir los consentimientos individuales. Cuando se hacen encuestas que no incluyen aspectos sensibles o confidenciales, se asume que la autorización puede ser verbal y no se hace necesaria la firma de un consentimiento informado.

Algunas investigaciones se consideran ***sin riesgo***: **las que emplean técnicas que no conllevan** riesgos, o tan solo recogen información ya existente y que no tiene carácter sensible o confidencial. Incluyen estudios basados en información recolectada previamente y donde no se hace ninguna intervención o modificación intencionada de variables biológicas, fisiológicas, psicológicas o sociales de los individuos participantes[13,15].

Pertenecen a este grupo de las investigaciones sin riesgo las revisiones de historias clínicas o de bases de datos, y las entrevistas o cuestionarios en los cuales no se tratan aspectos sensibles de la conducta de los individuos, y donde estos no son identificados personalmente.

EJEMPLO

Un estudio observacional descriptivo, que revisa historias clínicas para describir las complicaciones renales encontradas en pacientes con lupus eritematoso sistémico; o un estudio que busca asociación entre dislipidemia en hipertensión arterial en hombres con y sin obesidad.

De forma similar, existen investigaciones con ***riesgo mínimo***, **definido como aquel que** no es más probable ni de magnitud mayor que el asociado a procedimientos clínicos o psicológicos de rutina, como hacer un examen clínico convencional, recolectar excretas, obtener líquido amniótico o la placenta durante el parto, obtener dientes o cálculos extraídos por indicación terapéutica o extraer sangre por punción venosa en adultos, entre otros[15]. **En estas** circunstancias, el investigador puede solicitar al comité de investigaciones correspondiente la omisión de algunos elementos del consentimiento mencionados en la **tabla 36.1**[16].

EJEMPLO

Se busca conocer la asociación existente entre enfermedad periodontal severa (periodontitis) y elevación de la proteína C reactiva ultrasensible (PCRUS) en una población adulta, en la que se comparan los valores de PCRUS entre quienes tienen periodontitis y quienes no tienen enfermedad periodontal. El estudio requiere que se haga una toma de sangre venosa para la medición de PCRUS, **examen que no se habría** hecho si no fuera por el estudio.

Se quiere conocer el rendimiento diagnóstico de una nueva modalidad de evaluación fisiológica cardiovascular, las mediciones de presión arterial central y de velocidades de onda de pulso aórtica, para lo que se requiere una medición no invasiva de estos parámetros, y la presencia de hipertensión arterial resistente, determinada por la historia clínica del paciente. En ninguno de los dos exámenes hay invasión, ni riesgos para el paciente.

Finalmente, se habla de investigación con *riesgo superior al mínimo*, **estudios en los cuales** las probabilidades de afectar al sujeto de investigación son significativas, como en el caso de

Tabla 36.1 Requerimientos mínimos del documento de consentimiento informado (CI) —Se exponen los requerimientos mínimos según la normativa, así como otras recomendaciones generales y las relacionadas con muestras biológicas—

Requerimientos mínimos del CI según la Resolución 8430 de 1993
Justificación y objetivos de la investigación.
Procedimientos del estudio y su propósito, incluyendo la identificación de los procedimientos que son experimentales.
Las molestias o los riesgos esperados durante el desarrollo del estudio.
Los beneficios que puedan obtenerse por la participación en el estudio.
Los procedimientos alternativos que podrían ser ventajosos para el sujeto.
La garantía de recibir respuesta a cualquier pregunta y aclaración a cualquier duda acerca de los procedimientos, los riesgos, los beneficios y otros asuntos relacionados con la investigación y el tratamiento del sujeto.
La libertad de retirar su consentimiento en cualquier momento y dejar de participar en el estudio, sin que por ello se creen perjuicios para continuar con su cuidado y su 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 proporcionarle información actualizada obtenida durante el estudio, aunque esta pueda afectar la voluntad del sujeto para seguir participando.
La disponibilidad de tratamiento médico y la indemnización a la que legalmente tendría derecho, por parte de la institución responsable de la investigación, en caso de daños que le afecten directamente, causados por la investigación.
En caso de que haya gastos adicionales, estos serán cubiertos por el presupuesto de la investigación o de la institución responsable de ella.
Requerimientos generales del CI
Duración esperada de la participación del sujeto (tiempo de participación del sujeto en el estudio, incluidos seguimientos posteriores a la intervención).
Además de la explicación pormenorizada de las intervenciones y los tratamientos que tendrá el estudio, debe explicarse claramente el proceso de asignación aleatoria al grupo placebo o a uno de los tratamientos.
Cuando sea apropiado, se debe explicar con claridad que un tratamiento o un procedimiento en particular pueden poner en peligro al embrión o al feto, aunque la precisión y la naturaleza de dichos riesgos se desconozcan.
• Descripción completa de los aspectos de seguridad y custodia de los datos. • Explicación del proceso de contacto con la persona del comité que responderá las preguntas pertinentes, sobre los derechos del sujeto y sobre dudas acerca del estudio, además del proceso por seguir si se presenta algún daño derivado de la participación en el estudio.
Notificar que puede haber circunstancias donde la participación del sujeto puede ser terminada por el investigador, sin consentimiento por parte del sujeto.
Aclaración de que nuevos y significativos hallazgos encontrados a lo largo del estudio pueden relacionarse con la voluntad del sujeto de continuar o no en el estudio.
Número aproximado de sujetos que estarán involucrados en el estudio.
Aspectos especiales del CI respecto a muestras biológicas
Describir el proceso de anonimización de las muestras.
Darle al sujeto la opción de que la muestra sea desechada.
Describir el proceso por el cual el sujeto puede solicitar que se retire y se deseche la muestra.
Dar claridad sobre los análisis que se llevarán a cabo (por ejemplo, sin son análisis relacionados con el estudio a escala molecular de la enfermedad, o si se trata de análisis relacionados con el diagnóstico de una enfermedad de origen genético).
Si se trata de análisis relacionados con potenciales diagnósticos presintomáticos, debe haber total claridad sobre los procesos de informe de resultados (si existirán o no), y dar la opción de no conocer resultados, explicar los procesos de asesoría, y si los resultados pueden usarse o no para el diagnóstico en familiares.
Describir si los resultados de los análisis tendrán algún beneficio para los sujetos o para sus familiares.
Aclarar los procedimientos en los menores de edad que puedan ser sujeto de análisis, pero sean asintomáticos de la patología en estudio.
Aclarar, si es el caso, que las muestras serán manejadas en laboratorios clínicos ubicados en otro país.
Especificar si las muestras y los datos clínicos asociados serán almacenados para estudios posteriores, y dar la opción a los sujetos de determinar si ellos autorizan el análisis solo en el estudio presente o en futuros estudios.
Especificar las instituciones involucradas en la investigación.

Fuente: Elaboración propia.

estudios que necesitan exámenes radiológicos, el uso de medicamentos, el estudio de dispositivos nuevos, o de intervenciones quirúrgicas, técnicas invasivas, los que utilicen placebos o los que utilicen aleatorización para la asignación a la intervención.

EJEMPLO

Se busca evaluar la efectividad terapéutica de una nueva modalidad terapéutica: el uso de anticuerpos monoclonales contra la PCSK-9, una proproteína de importancia en la hipercolesterolemia familiar. Los participantes serán aleatorizados (criterio de riesgo superior al mínimo) a recibir el medicamento nuevo (otro criterio) o a recibir placebo (otro placebo).

En investigaciones de tipo epidemiológico, que usualmente solo conllevan un riesgo mínimo, es usual que el investigador solicite la aprobación y la cooperación de las autoridades locales o nacionales responsables de la salud pública de la población.

En el caso de comunidades con cuerpos de decisión colegiados, es aconsejable buscar el consentimiento de la comunidad por medio de sus representantes legítimos. Si la investigación implica un contacto personal entre los investigadores e individuos de la comunidad, se debe solicitar también su consentimiento a cada sujeto. Por otra parte, si la participación de cada individuo se hace, fundamentalmente, como miembro de un grupo poblacional, puede ser aceptable no obtener el consentimiento de cada sujeto, pero es indispensable que quien dé el consentimiento grupal tenga la autoridad investida para hacerlo así, o esté reconocido por la comunidad como su representante.

Existen circunstancias en las cuales, para garantizar la validez de la investigación, el investigador puede deliberadamente ocultar parte de la información que se le da al participante en el momento de solicitar el consentimiento. Esto se conoce como ***restricción de información***. En la forma más común de ello, el investigador se reserva parte de la información respecto a ciertos procedimientos del estudio (por ejemplo, no se les informa a los participantes que el propósito de ciertas pruebas es medir su adherencia a los fármacos empleados en el estudio).

Con menor frecuencia se le puede dar al sujeto información incorrecta, como puede ocurrir en investigaciones donde el investigador pretende ser un paciente, para así evaluar las actitudes y el comportamiento de los profesionales de la salud que lo atienden. Se habla de ***engaño justificado*** en esos casos en los que la información que se les da a los participantes no es completa o no es correcta, por razones válidas. En cualquier caso, estos procedimientos deben ser explícitamente autorizados por el comité de investigaciones, el cual debe revisar si hay justificación para el cambio de información o para la omisión. Si hay justificación, el procedimiento podrá ser aprobado sin reparos bioéticos.

EJEMPLO

Se busca evaluar el impacto, en cuanto al número de consultas, de complicaciones o de hospitalizaciones en sujetos diabéticos a quienes se les imparte un programa de educación en su enfermedad. Si se les invita a participar en el estudio y si se les explica que se trata de evaluar si haber recibido educación sobre diabetes disminuye el número de consultas o de hospitalizaciones, esta información puede influir en la conducta de los sujetos, que pueden abstenerse de consultar para tener un resultado "bueno", acorde con los objetivos del estudio. En este caso hay razones para un *engaño justificado*, y se les puede explicar que el estudio va a evaluar cambios en una variable biológica (por ejemplo, microalbuminuria) durante un año. Se les explica que, como parte del estudio, y dado que tendrán que asistir a consulta una vez por mes para la toma de las muestras, se les ofrecerá un programa de educación en varios aspectos relacionados con la diabetes, como principios de cocina para diabéticos o educación general sobre la enfermedad. De esta manera, un grupo recibirá la intervención que se va a evaluar (educación), el otro grupo recibirá la intervención de control (principios de cocina) y ninguno de los participantes sabrá que el desenlace será medir el número de visitas extraordinarias, de complicaciones y de hospitalizaciones.

Finalmente, según lo descrito, el sujeto debe tomar la decisión respecto a su participación en el proyecto, sin coacciones ni intimidaciones. Esto no descarta la posibilidad de ofrecerle el reembolso de algunos gastos asociados a la participación en el estudio (transporte, honorarios no percibidos), así como gratificaciones razonables por las molestias asociadas a su participación. Si se hacen en dinero, dichos pagos deben guardar proporción con el esfuerzo y el riesgo, y no deben ser tan elevados como para que lleven al individuo a aceptar riesgos indebidos

o a una participación forzada, en contra de su percepción global[4].

El consentimiento es el cierre del marco que comprende los principios éticos, y que permite acciones prácticas ante a las fallas que se susciten en el desarrollo investigativo.

Por otra parte, hay consenso en el mundo sobre los puntos mínimos que deben incluirse de forma explícita en el consentimiento **(tabla 36.1)**[13,14].

MARCO LEGAL DE LA INVESTIGACIÓN CLÍNICA

Los aspectos éticos de la investigación no pueden desprenderse de los aspectos legales, que, en la mayoría de los casos responden a la formalización de acuerdos universales entre investigadores y filósofos alrededor de ciertos principios que merecen una representación en la ley o en la normatividad local, por ejemplo, la Resolución 8430 de 1993, por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud en Colombia[15]. Por otra parte, existe una serie de documentos que, si bien no son reconocidos estrictamente dentro de la legislación de los países, sí son referentes frente a la reflexión y el quehacer de la ética de la investigación, como por ejemplo, el mencionado Informe Belmont o los documentos de CIOMS. Ambas fuentes de información, las normas y las declaraciones universales son referentes obligados para el análisis ético de la investigación científica.

A continuación se relacionan las principales leyes y normas que rigen la investigación científica en Colombia; así mismo, se listan declaraciones universales que son relevantes en el ámbito de los comités.

Normatividad nacional vigente en relación con la investigación

En primer lugar el quehacer investigativo se rige por la Ley 23 de 1981, o Ley de Ética Médica[17].

En ella se estableció que la comunidad médica nacional debe seguir las recomendaciones de la Asociación Médica Mundial (AMM) en relación con la "investigación biomédica en general y con la investigación terapéutica en humanos; aplicación de nuevas tecnologías, tanto con fines terapéuticos, como es el caso de algunos tipos de cirugía cardiovascular y psicocirugía y experimentación en psiquiatría y psicología médica y utilización de placebos". Es decir, la Ley de Ética Médica toma como referente la declaración de Helsinki, de la AMM, como el marco general de los aspectos éticos de la investigación científica, de tal forma que a partir de dicha declaración y de sus respectivas actualizaciones se genera el marco legal general de la investigación científica en seres humanos en el país.

Después de la Ley 23 de 1981[15], se promulgó la Resolución 8430 de 1993, que establece las normas que enmarcan propiamente la investigación científica; la norma toca diversos aspectos, que incluyen la valoración del riesgo para los sujetos en las investigaciones y los parámetros del CI.

Posteriormente, y en consecuencia con el desarrollo de la investigación clínica en el país, en su momento, el Ministerio de la Protección Social emitió la Resolución 2378 de 2008[18], por la cual se adoptan las directrices de Buenas Prácticas Clínicas (BPC), desarrolladas, a su vez, por la Conferencia Internacional de Armonización (ICH, por las sigla en inglés de International Conference for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use), y que se han convertido en el estándar de funcionamiento de la investigación clínica en el mundo. La Resolución 2378 incluye los parámetros de funcionamiento y conformación de un comité de investigación y ética, y los de las unidades de investigación y farmacia de un centro dedicado a la investigación científica, así como los parámetros de calidad y desempeño de los equipos de investigación. Aunque las *Guías de Buena Práctica Clínica* fueron diseñadas y adoptadas para la investigación con medicamentos, sus recomendaciones, aceptadas por la Resolución 2378, han sido adoptadas como un marco reglamentario para la investigación clínica, aun si no involucra medicamentos. El Instituto Nacional de Vigilancia de Alimentos y Medicamentos (Invima) ha precisado los alcances de la Resolución 2378 a través de diversas circulares que aclaran la aplicación de la norma respecto al

funcionamiento de los comités, al desempeño de los investigadores y al reporte de eventos adversos[19].

Ambas resoluciones (8430 de 1993 y 2378 de 2008) constituyen la normatividad que regula, principalmente desde el punto de vista legal, la investigación científica del país; sin embargo, vale la pena mencionar otras normas dignas de tenerse en cuenta, según el ámbito investigativo y el manejo de los datos recolectados y los derivados de la investigación.

Como ejemplo de lo anterior, las investigaciones que se relacionen con cadáveres y componentes anatómicos deben ceñirse a la Resolución 485 de 2002, por medio de la cual se reglamenta el procedimiento para la entrega de cadáveres y de componentes anatómicos con fines de docencia e investigación[20]. Por otra parte, las investigaciones documentales que involucren la revisión o la extracción de datos de historias clínicas deben contemplar la normatividad vigente descrita en la Resolución 1995 de 1999, por la cual se establecen normas para el manejo de la historia clínica[21].

Por último, se debe tener en cuenta la normatividad vigente relacionada con datos personales e información sensible; específicamente, la Ley Estatutaria 1581 de 2012, en la que se dictan disposiciones generales para la protección de datos personales[22].

Como ya se mencionó, el marco legal comprende, estrictamente, la normatividad y las leyes locales; sin embargo, el referente universal es relevante para el desarrollo y la evaluación de los protocolos de investigación. En la **tabla 36.2** se resumen las principales declaraciones y documentos universales de relevancia en la ética de la investigación, y se listan las normas locales al respecto.

INVESTIGACIÓN EN SUJETOS ESPECIALES

Ocasionalmente, es necesario adelantar estudios con la participación de sujetos que tienen limitada su capacidad para dar un verdadero CI. Esto incluye a menores de edad, a individuos con alteraciones mentales o del comportamiento, a subordinados (empleados, estudiantes o militares) o a personas privadas de su libertad.



En los casos mencionados, teniendo en cuenta el marco ético, deben cumplirse varias condiciones indispensables; primera y más importante: solo habrá justificación para llevar a cabo dichos estudios cuando no sea posible obtener la información relevante en otras poblaciones no limitadas.

Por ejemplo, no se hará investigación con menores de edad, a menos que el estudio no pueda hacerse en otras poblaciones. Segunda, el propósito del estudio debe ser la obtención de conocimiento fundamental para las necesidades de salud de las personas con limitaciones a quienes se pretende estudiar. Tercera, el riesgo que representen las intervenciones no relacionadas con beneficio directo debe ser bajo, y ser proporcional a la importancia del conocimiento que se obtendrá, y las intervenciones de las cuales se espera que produzcan beneficio deben ser, al menos, potencialmente tan ventajosas como cualquier otra alternativa disponible. Finalmente, debe obtenerse el consentimiento de un representante legal o de otra persona debidamente autorizada, y quien debe cumplir a cabalidad alguna de las dos descripciones anteriores.

Ni la mayor parte de la información respecto a las enfermedades de los niños ni su tratamiento pueden obtenerse en poblaciones que no tienen limitaciones para dar su CI: por ejemplo, en adultos. Las diferencias biológicas y de comportamiento propias de cada edad limitarían la aplicabilidad de la información obtenida de los menores —en estas circunstancias, es necesario incluir a niños como sujetos del estudio—. Para que esto sea aceptado, el investigador debe demostrar la imposibilidad de obtener el mismo tipo de conocimiento estudiando a poblaciones menos vulnerables, así como demostrar que el propósito de la investigación es beneficiar predominantemente a los niños. Lo anterior implica que no se debe incluir a niños en una investigación si la información puede obtenerse estudiando solo a adultos.

Adicionalmente, se deberá obtener CI tanto de los padres o de los representantes legales como del propio niño, a quien debe explicársele, en términos apropiados, el propósito del estudio. Una negativa por parte del menor deberá respetarse aún en los casos donde el adulto responsable haya aceptado participar[13,14,24-26].

Tabla 36.2 Normas y declaraciones de interés para el análisis ético de los protocolos de investigación

Normas y declaraciones universales	
Norma	**Organización**
Código de Núremberg (1947)[23].	Tribunal Internacional de Núremberg.
Declaración de Helsinki (1964)[2].	AMM.
Reporte Belmont (1979)[4].	Comisión Nacional para la Protección de Sujetos Humanos de Investigación Biomédica y de Comportamiento.
Guía de Buenas Prácticas Clínicas (1996)[6].	ICH.
Pautas Éticas Internacionales para la Investigación y Experimentación Biomédica en Seres Humanos, de la OMS (2003)[14].	CIOMS.
Pautas éticas para la evaluación ética de los estudios epidemiológicos (1993)[13].	CIOMS.
Leyes y normas locales	
Ley/norma	**Institución**
Ley 23 de 1992 (Ley de Ética Médica)[17].	Congreso de la República de Colombia.
Resolución 8430 de 1993 (normas de investigación científica[15].	Ministerio de Salud de Colombia.
Resolución 2378 de 2008 (adopción de las BPC)[18].	Ministerio de la Protección Social.
Resolución 485 de 2002 (investigación con cadáveres y componentes anatómicos)[20].	Instituto Nacional de Medicina Legal y Ciencias Forenses.
Resolución 1995 de 1999 (historias clínicas)[21].	Ministerio de Salud de Colombia.
Ley Estatutaria 1581 de 2012 (*habeas data*).	Congreso de la República de Colombia.

Fuente: Elaboración propia.

Las mismas consideraciones se aplican a sujetos con enfermedades mentales o incapaces de dar su consentimiento por sí mismos. En el caso de individuos institucionalizados o sin representante legal, debe obtenerse la autorización de una autoridad competente.

Un aspecto interesante y poco estudiado de la investigación con poblaciones especiales (en particular, con niños) es el relacionado con el obstáculo creciente que significa la negativa, cada vez más habitual, de los padres a autorizar la participación de sus hijos. Esta ha sido considerada una de las causas más importantes de dificultades en la investigación después de los problemas de financiamiento[27].

La calidad del consentimiento emitido por sujetos que tienen algún grado de subordinación a los investigadores o a los patrocinadores del estudio también demanda una muy cuidadosa consideración. Esto incluye a estudiantes y a empleados de la institución investigadora o a sujetos que pueden verse forzados a aceptar obedeciendo órdenes (miembros de instituciones altamente jerárquicas, como las Fuerzas Armadas [Policía y Ejército] o grupos o instituciones religiosas) o por temor a represalias o a malas calificaciones. Dadas las relaciones entre ellos y los investigadores, se pueden presentar dudas respecto a la verdadera capacidad del sujeto para negarse a participar. Además, su proximidad con los investigadores o con sus superiores disciplinarios puede llevarlos a ser llamados a participar en investigaciones más a menudo que otras personas, lo cual rompe el principio de selección justa de los participantes, ya mencionado. Otros grupos vulnerables son los residentes en hogares de cuidados especiales, los pacientes en situaciones críticas o de emergencia, los desempleados o los beneficiarios de programas de asistencia social, las minorías étnicas, los refugiados, los nómadas o las personas con enfermedades incurables. Todos estos casos requieren una muy cuidadosa evaluación por los comités de ética.

Como un aspecto importante de la evaluación de vulnerabilidad, la carencia de recursos económicos o la extrema necesidad deben ser considerados marcadores de vulnerabilidad, y a las poblaciones en estas condiciones se las debe tratar con las mismas consideraciones especiales que se ofrecen a los otros grupos mencionados.

Parece obvio que las poblaciones en desventaja no deberían ser parte de proyectos de investigación. Sin embargo, hay un derecho inalienable a ser potencialmente beneficiado por un medicamento o por una intervención nueva mediante un proyecto de investigación, o, simplemente, a participar en un proyecto que, pese a no ofrecer beneficio personal, le permita a la persona colaborar.

El derecho en mención, protegido por las guías bioéticas, evita que las personas en desventaja sean excluidas injustamente de los proyectos de investigación. Por ejemplo, la mayor parte de las compañías farmacéuticas adelantan los estudios para el desarrollo de nuevos medicamentos en adultos; como consecuencia, la información obtenida en dichas investigaciones no permite conocer la efectividad o la seguridad de los mismos productos en otras poblaciones, lo cual puede implicar su empleo clínico sin la adecuada información. Se mostró en un estudio, por ejemplo, que en algunos países europeos cerca de la mitad de las prescripciones en niños hospitalizados correspondieron a drogas sin licencia para uso en menores, o a prescripciones fuera de los términos de la licencia del producto[28]. La misma situación se ha descrito para mujeres embarazadas o en mujeres en edad fértil.

Por otro lado, hay bastante controversia respecto a la participación de prisioneros en proyectos de investigación en salud. Quienes están a favor mencionan que ellos pueden ser particularmente atractivos, debido a las condiciones estándar en las cuales se ven forzados a vivir y a la capacidad de garantizar seguimiento por largos periodos. También se menciona que esto puede ayudarles a sentirse valiosos, a librarse del tedio propio de su encierro y a obtener algunos ingresos económicos. Los opositores claman que los prisioneros no pueden dar un verdadero CI. No obstante, hay acuerdo en que los prisioneros que sufren enfermedades graves no deben verse privados de los potenciales beneficios de participar en proyectos de investigación que evalúen nuevas formas de tratamiento con potencial beneficio para sus dolencias.

INVESTIGACIÓN EN COMUNIDADES MENOS DESARROLLADAS

Un aspecto sobre el cual hay gran debate en la literatura médica y en los comités de investigación y ética es la conducción de proyectos en países o comunidades con pobre desarrollo. Existe acuerdo en varios aspectos. En primer lugar, es claro que el investigador de una comunidad o un país desarrollado que adelante proyectos en uno menos desarrollado debe respetar los estándares éticos de su propio país y las características culturales de la sociedad donde se adelanta la investigación.



Como norma, se ha propuesto que en caso de haber diferencias en la legislación entre el país del investigador y el país donde se obtendrán los datos, debe respetarse siempre la legislación más estricta y exigente de las dos, o la de los países involucrados.

EJEMPLO

Se va a conducir en varios países de América Latina un estudio, originado en un país europeo, de intervención en patrones de alimentación y ejercicio, en un experimento clínico aleatorizado (ECA). Entre los criterios de inclusión está que el participante sea adulto. La legislación de los países participantes en América Latina define a los menores de edad como los menores de 16 años, en tanto que en el país europeo patrocinador la legislación es diferente, y el límite de edad es de 18 años. El estudio debe adoptar como criterio de definición el más exigente; en el caso planteado, el de 16 años, aunque el protocolo haya sido originado en otro país y tenga criterios de inclusión preestablecidos. Independientemente de cuál sea el país de origen del protocolo, del investigador, de los sujetos o del patrocinador, debe tomarse como definición la que signifique el criterio más estricto.

También hay consenso en que estos proyectos de investigación deben dar información que sea de utilidad para las necesidades del país donde se adelanta el estudio. Otro punto de acuerdo es que las diferentes fases de los estudios que desarrollan fármacos (I-V) deben conducirse simultáneamente en el país patrocinador y en el anfitrión[13,15].

La controversia gira alrededor de la investigación sobre fármacos o tecnologías costosas que quizás no estén al alcance del país anfitrión; en especial, respecto a las terapias empleadas en el grupo control, el cual, por las limitaciones económicas del país donde se adelanta el estudio, usualmente no incluye tratamientos que se consideran apropiados en lugares más desarrollados[29].

Quienes se oponen a estas prácticas señalan que solo es ético dejar de usar un tratamiento o emplear placebo en el grupo control para evaluar una nueva intervención cuando no existe un tratamiento efectivo conocido. Si dicho tratamiento existe, el investigador les debe ofrecer a los sujetos del grupo control la terapia que ya ha demostrado ser efectiva.

La otra línea de argumentación menciona que en las circunstancias descritas no se está privando a los individuos participantes de ninguna intervención, pues si bien existen tratamientos que pueden ser efectivos, estos no se hallan disponibles, debido a las limitaciones económicas o de desarrollo propias del mundo real. Desde este punto de vista, los sujetos participantes estarían recibiendo cuidados similares a los de sus conciudadanos que no están participando en el estudio. Aún no hay acuerdo definitivo sobre este dilema.

ÉTICA EN LOS EXPERIMENTOS CLÍNICOS

Se define como investigación:

> [el] conjunto de actividades diseñadas para desarrollar o contribuir al conocimiento generalizable, que consiste en teorías, principios o relaciones, o a la acumulación de información en la que están basadas, que pueden ser corroborados por métodos científicos aceptados de observación e inferencia [...] Se diferencia la investigación biomédica (en referencia a su relación con salud) en aquella que se hace simultáneamente con atención médica (investigación clínica) y la que se hace con pacientes o con otros sujetos o con datos que les pertenecen, solamente para contribuir al conocimiento generalizable (investigación biomédica no clínica)[14].

La participación de los médicos como investigadores en los estudios clínicos (particularmente, en ECA) puede implicar un conflicto aparente para el médico investigador, quien puede sentirse forzado a modificar sus obligaciones morales con el paciente al hacer un estudio en el que el azar determina el tratamiento que el sujeto recibirá. Teóricamente, ello implicaría el sacrificio del juramento de ofrecerles lo mejor a los pacientes que tiene a su cuidado[24,25]. Este dilema aparente, pero preocupante en grado sumo para los clínicos, desaparece por completo al considerar la evidencia que apoya el uso de los tratamientos que se pretenden comparar en la investigación.

De esta manera, no tener un conocimiento cierto sobre la superioridad de un tratamiento justifica la aleatorización de las intervenciones y convierte el estudio en una necesidad. En algunos casos podría considerarse, incluso, que no responder la pregunta es una falta ética, si responderla está al alcance del clínico investigador. Si es así, el empleo del azar para establecer el tratamiento que recibe cada participante no plantea dilemas éticos, pues no se está privando a los participantes de una forma de tratamiento que se reconozca como superior[30,31].

Con todo, se acepta que el mejor instrumento de investigación disponible para evaluar la efectividad de una intervención es el ECA, pues protege a los clínicos y a los pacientes contra los riesgos de intervenciones no efectivas o peligrosas que podrían parecer más efectivas que lo que en realidad son si fueran estudiadas con diseños de investigación de inferior calidad[32]. Aún más, la aplicación estricta de dicho argumento implica que lo ético es adelantar el mejor estudio posible; o sea, un ECA.

El ***equilibrio clínico*** (del inglés *clinical equipoise*), también llamado ***principio de indeterminación clínica*** o ***de incertidumbre terapéutica***, es uno de los principios que proporcionan las bases éticas sobre las cuales se desarrolla la investigación clínica que involucra la asignación aleatoria de sujetos a diferentes grupos de tratamiento en un experimento clínico[30]. Este principio surge de los desafíos de la investigación en el ámbito asistencial, que, a su vez, se contraponen con las características propias del experimento, las cuales disminuyen los errores de medición y promueven las acciones para evitar los sesgos e identificar factores de confusión. En la **tabla 36.3**, se resumen los desafíos que se contraponen a las características propias del experimento.

Ante el contraste entre los desafíos que plantea el experimento clínico a la atención médica y a sus ventajas como herramienta de investigación, el ***principio de incertidumbre clínica*** intenta dirimir la diferencia entre ambas posiciones y pretende fundamentar que la asignación aleatoria sí es ética; la noción de equilibrio significa que existe una incertidumbre genuina, por parte de la comunidad médica, sobre cuál tratamiento será benéfico en alguna enfermedad.

Tabla 36.3 Desafíos de la experimentación en el ámbito asistencial

Desafíos de la experimentación
En un experimento clínico en el que se comparan dos (o más) tratamientos distintos, hay el riesgo de que uno de los grupos o brazos de tratamiento reciba un tratamiento subóptimo.
La asignación aleatoria es el proceso que determina el tratamiento para cada sujeto, y así se impide que sea el médico, de acuerdo con su criterio clínico, quien lo decida.
El protocolo define el perfil terapéutico de la intervención: dosis, frecuencia, horarios y vías de administración de la medicación, o la intervención en estudio. Estos aspectos limitan la instauración de manejos personalizados del tratamiento farmacológico o no farmacológico, según el criterio médico.
El protocolo define cuáles tratamientos concomitantes están permitidos o excluidos, bien sea por razones de seguridad del sujeto (interacciones con el medicamento en estudio), o bien, para asegurar la evaluación no sesgada del tratamiento en estudio.
En la mayoría de los casos, tanto las evaluaciones de eficacia del tratamiento experimental como las evaluaciones de seguridad requieren diversos estudios paraclínicos, radiografías o pruebas de laboratorio, que son de utilidad para las evaluaciones experimentales, pero no son necesarios en la rutina de atención del paciente, lo cual implica diversos riesgos y gastos adicionales.
Algunos diseños de experimentación clínica requieren *periodos de lavado* (suspensión de un tratamiento) antes de iniciar el estudio. Esto implica que el paciente (sujeto de investigación) no está "cubierto" por el tratamiento que venía recibiendo, y eso, ocasionalmente, lo pone en riesgo de recaída o de empeoramiento de su condición.
El médico asume una posición moral ambigua frente a su paciente cuando le ofrece participar en un experimento clínico. La ambigüedad surge del papel que debe cumplir el médico: el de tratante frente al de investigador. Como tratante, su obligación es la búsqueda del máximo beneficio para su paciente; como investigador, su principal interés es la búsqueda de conocimiento, aun cuando dicho interés científico se basa en la búsqueda de conocimiento sobre potenciales tratamientos benéficos para el mismo paciente o para futuros pacientes.
La puesta en marcha de un experimento clínico obliga al médico-investigador a preguntarse si la investigación científica justifica moralmente definir el manejo de la enfermedad de un paciente, no ya según el criterio médico y el tratamiento individualizado, sino según el manejo de la enfermedad que se propone en el protocolo experimental.
Características del experimento clínico
El uso de controles concurrentes en el experimento permite eliminar las diferencias entre los grupos debidas a tendencias temporales, y permite la comparación directa entre los brazos de tratamiento.
La aleatorización reduce la posibilidad de que los grupos difieran al inicio del estudio.
El empleo de diversas pruebas estadísticas, además de la definición y el análisis de las variables genera resultados de importancia tanto estadística como clínica, con diversos grados de confianza.
El uso adecuado del cálculo del tamaño de muestra implica que los grupos son representativos de la población en estudio.
El enmascaramiento reduce la posibilidad de que los investigadores interactúen con los sujetos participantes de tal forma que puedan sesgar las observaciones, y reduce la posibilidad de que los sujetos reporten síntomas o se comporten sesgadamente al conocer el brazo de tratamiento en el que se encuentran. Permite, además, la distinción de los efectos fisiológicos, y así posibilita diferenciar el efecto del placebo.
El uso del placebo facilita el enmascaramiento cuando las intervenciones difieren en formas observables por los sujetos.

Fuente: Elaboración propia.



Al existir esa incertidumbre legítima, sincera y honesta, la comparación de múltiples métodos de tratamiento (en un experimento clínico) sería también legítima, pues el médico desconoce la superioridad de un manejo sobre otros.

En tal situación, el médico realmente no tiene una opinión formada sobre un tratamiento, y la asignación aleatoria, necesaria para fundamentar científicamente futuras decisiones terapéuticas, es éticamente aceptable, por cuanto el médico no abandona la atención personalizada del sujeto, sino que deja al azar la asignación del tratamiento para cada caso. En el concepto de equilibrio, la incertidumbre honesta es de carácter colectivo, pues define al desacuerdo como un desacuerdo profesional entre expertos clínicos (la comunidad de médicos). Y ni la ausencia de consenso entre profesionales ni la presencia de "corazonadas" o de preferencias personales

por parte de los médicos frente a un tratamiento son barreras morales para reclutar en el experimento clínico.

En otras palabras, dadas las condiciones de incertidumbre, los médicos deben ver los tratamientos ofrecidos por el experimento clínico como equivalentes (***equipoise***) a los ofrecidos por la atención clínica personalizada, una equivalencia en la perspectiva de la duda sobre la eficacia de un manejo y en la perspectiva del azar.

En términos más coloquiales, lo anterior equivale a decir que si la comunidad médica no sabe si el "tratamiento ***A*" es mejor que el "tratamiento *B*", entonces dar uno u otro tratamiento** al azar equivale a elegir ***A*** o ***B*** en el ámbito clínico, solo que el azar embebido dentro del experimento ofrecerá la respuesta que resuelva la duda. La condición para incluir a un paciente es el estado de duda, o ***incertidumbre***.

Un aspecto adicional, que en ocasiones plantea dilemas éticos en experimentos controlados, es el empleo de placebo como terapia de control para establecer la efectividad de una nueva intervención. El empleo de placebo solo será éticamente aceptable si no está disponible una forma de tratamiento cuyo valor ya haya sido demostrado. Si dicha forma de terapia existe, lo apropiado es ofrecérsela a los individuos que conforman el grupo control[16].

Por otra parte, si el uso de placebo es apropiado, su empleo puede llevar a que el diseño del experimento sea más adecuado (pues facilita el enmascaramiento a todas las partes involucradas); es decir, implica que el estudio tendrá más probabilidades de dar una respuesta válida. En otras palabras, en algunas circunstancias lo que puede cuestionarse desde el punto de vista ético es no emplear placebo cuando su uso está indicado por razones metodológicas.

Como norma fundamental, entonces, el placebo como única alternativa solo podrá ser utilizado cuando no haya un tratamiento reconocido para la condición de estudio.

Pero, con alguna frecuencia el placebo puede ser una adición a un tratamiento establecido de base, para facilitar la comparación enmascarada con otro medicamento. Si se tiene a un paciente con epilepsia no controlada con el medicamento de base y se quiere estudiar la adición de un segundo antiepiléptico, el placebo podrá utilizarse. Así, todos los pacientes recibirán el tratamiento de base (ácido valproico, por ejemplo), y la mitad recibirán, por azar, placebo o lamotrigina.

Igualmente importante es la consideración de la calidad de medición, para lo cual es vital el enmascaramiento (cegamiento) del tratamiento tanto al paciente (***estudio enmascarado***) **como a quien hace la evaluación (*estudio doble enmascarado*), o, incluso, a quien hace** el análisis (***estudio triple enmascarado***). **No** aplicar el enmascaramiento o, en algunos casos, no aplicar la restricción en información puede demeritar la calidad del estudio y llevar a que sea antiético por poner en riesgo la validez de los resultados.

Debe mencionarse que se prefiere en la actualidad el uso del término ***enmascarado*, por** las connotaciones de discapacidad del término antiguo (***ciego***).

En los casos cuyos métodos no sean directamente comparables (por ejemplo, dos medicamentos con presentaciones diferentes, como tableta y cápsula) pueden emplearse dos soluciones: 1) el ***doble enmascaramiento*** **(los** pacientes de un grupo reciben una tableta con medicamento y una cápsula con placebo, y los sujetos del otro grupo reciben tableta con placebo y cápsula con medicamento), o 2) la selección de variables "duras" (las que son objetivas o en las que hay poca influencia externa) en contraposición con variables "blandas" (en las que el método de evaluación es subjetivo, o la medición puede ser fácilmente modificable por circunstancias externas o internas). No aplicar los métodos para la correcta interpretación de los resultados de las mediciones atenta contra la calidad del estudio (***validez interna***) **y es una clara falta a la ética.**

INVESTIGACIÓN, ÉTICA Y DATOS GENÉTICOS

La investigación clínica recolecta diversos datos que definen las variables del estudio, y dentro de estos datos, la información genética ha cobrado cada vez mayor importancia en la reflexión ética, debido a su relevancia como determinante de la aparición y el curso de la enfermedad, así como de la respuesta terapéutica y de pronóstico a largo plazo. Además de ser factores de riesgo para diversas condiciones clínicas y respuestas farmacológicas, los datos genéticos implican la herencia, los riesgos familiares y la posibilidad de establecer predicciones para

sujetos asintomáticos, o, eventualmente, para futuras generaciones[33].

El manejo de esa información por parte de los investigadores implica un manejo especial de la información que requiere, en la mayoría de los casos, un CI especial, dedicado a recolectar información genética. Esta condición especial de los experimentos clínicos da para que los subestudios dedicados al estudio de la *farmacogenómica*, o estudios de marcadores moleculares de respuesta a la intervención en estudio, tengan un espacio de discusión con los sujetos sobre las implicaciones de tomar muestras para fines de estudios sobe los genes. Este espacio se hace necesario, dado que la información genética implica situaciones en las que el manejo de los resultados puede repercutir en áreas tan disímiles como el aseguramiento en salud o la detección de enfermedades para las cuales el estudio no fue diseñado[33,34].

En primer lugar, se deben definir los beneficios que para el sujeto se deriven del estudio de los genes, y esta información debe ser explícita en el consentimiento. En la mayoría de casos en los que se estudian determinantes genéticos y biológicos de las enfermedades prevalentes, se trata de marcadores polimórficos que, aun en conjunto, no implican la detección de otras enfermedades del individuo, ni implican que al obtener esa información se deba proceder a intervenciones diagnósticas o terapéuticas sobre el sujeto en quien se recolectan las muestras.

Lo anterior se trata, sin embargo, de datos relevantes para entender la respuesta a una intervención farmacológica o las características de una enfermedad; en esos casos, el beneficio para el individuo es mínimo o nulo, lo cual debe ser explícito en el proceso de consentimiento. Por otra parte, la recolección de dichas muestras no es obligatoria, y los sujetos pueden participar en el estudio de intervención sin tener que participar en subestudios que impliquen la toma de muestras para análisis genéticos. En ese sentido, ni los derechos del sujeto ni la selección justa de los participantes deben depender de la intención del sujeto de suministrar muestras para estudios genéticos.

Aun cuando la intención de los estudios de polimorfismos en experimentos clínicos no sea un ejercicio diagnóstico directo o indirecto, son un conjunto de información muy íntima de cada uno de los sujetos que participan en la investigación, y el investigador debe garantizar el manejo adecuado de las muestras, lo que incluye: la seguridad de las muestras y su trazabilidad; la anonimización de estas; que se les brinde la posibilidad a los sujetos de destruir las muestras, de saber si las muestras se destruirán en algún momento o de saber si permanecerán dentro de un biobanco; de aclarar si las muestras saldrán del país para análisis específicos; de conocer si se utilizarán como parte de otros estudios aparte del actual. Finalmente, el investigador debe establecer con claridad si los sujetos recibirán información, y cuál, sobre los resultados de los análisis.

Las implicaciones del estudio genético sobrepasan los alcances del presente capítulo, pero, teniendo en cuenta los principios y el marco ético general ya enunciado, los investigadores deben considerar que si el diseño metodológico implica el uso de marcadores genéticos, estos deben tener un tratamiento especial, que implica la mayor cantidad de información para el sujeto reclutado; también, que se explique el alcance de los estudios que se van a hacer y si esos resultados tendrán alguna importancia sobre su estado de salud o sobre la posibilidad de que esos mismos resultados tengan implicaciones para otros miembros de la familia, sean sintomáticos o no. En consecuencia, el equipo de investigación debe planear con anticipación la forma como se manejarán los datos en relación con diagnósticos y consejería respectiva, y asegurar la logística de manejo de las muestras biológicas.



La confidencialidad y la privacidad de los datos son un aspecto fundamental en el estudio de marcadores genéticos, debido a que el uso inadecuado de dichos marcadores podría suscitar discriminación o señalamientos, según ciertas condiciones biológicas.

Como, por ejemplo, la exclusión de un seguro médico si se tiene una susceptibilidad elevada a padecer un trastorno y esta susceptibilidad se descubre en estudio clínico.

La tecnología ha hecho posible que el estudio de los genomas de los sujetos involucrados en investigación sea tan rápido y eficiente como la recolección de datos de cualquier otra variable fisiológica, de modo que la reflexión en torno a las implicaciones éticas y al manejo de esta información se haya rezagado frente a la ca-

pacidad tecnológica actual. En consecuencia, la normatividad tampoco alcanza a estar al día con los adelantos y la velocidad del manejo de la información.

La reflexión ética pude valerse de principios universales en torno al genoma humano que sirvan como referente al respecto. Por ejemplo, en la Declaración de Inuyama, de 1990, se explican los beneficios de la investigación de orden genómico, y, al mismo tiempo, se advierte sobre el peligro del determinismo genético que ha rodeado las investigaciones biológicas[35]. En la forma más radical del determinismo genético, se expone que tanto las manifestaciones físicas como las conductuales están determinadas por los genes. Parecería más razonable que, si bien un gen aislado puede estar claramente asociado a un fenotipo, no todos los fenotipos están asociados a un gen aislado (a veces, con más de uno; y a veces están influidos por el ambiente). Ello explicaría que si bien una alteración genética como el síndrome de Down tiene una base perfectamente clara en la trisomía 21, el desarrollo conductual, por ejemplo, del individuo tiene una clarísima influencia del medio en el que él se desarrolle. No siempre la predisposición genética significará la aparición del trastorno, y no siempre que este aparezca tendrá las mismas condiciones en un porcentaje apreciable de los casos.

De igual forma, la reflexión ética en relación con datos genéticos y con su uso en investigación debe tener en cuenta la singularidad de dichos datos y las finalidades para las cuales serán empleados en un protocolo de investigación científica, tal como lo expresan la Declaración Universal sobre el Genoma Humano y los Derechos Humanos, y la Declaración Internacional sobre los Datos Genéticos, de la United Nations Educational, Scientific and Cultural Organization (UNESCO)[36,37].

Debería siempre estar perfectamente claro en el protocolo (y expresarlo así al potencial participante en un estudio durante el proceso de CI) cuáles son los objetivos específicos del estudio, y cuáles, los usos detallados que se les darán a las muestras que se toman, cuáles pruebas se harán, qué información se buscará y si se guardarán o no muestras, y cuál será el uso de estas en el futuro. Cada uno de dichos aspectos debe estar claro para el participante potencial, y él debe tener potestad para decidir sobre cada una de ellas.

Con alguna frecuencia se hace un proceso diferente de CI para la parte genética, y puede ser deseable que haya dos procesos diferentes y se firmen dos documentos de aceptación y de consentimiento: uno para los procedimientos generales del estudio y uno para las pruebas genéticas, pues podría ser que el individuo consienta su participación para el estudio general, pero no para el estudio genético.

CONSTITUCIÓN Y RESPONSABILIDADES DE LOS COMITÉS DE INVESTIGACIÓN Y ÉTICA

Todos los proyectos de investigación que impliquen la participación de sujetos humanos deben ser estudiados y aprobados por comités de revisión ética y científica, los cuales deben ser conformados como comités de investigación y ética, y deben ser independientes de los investigadores y de los patrocinadores de los estudios. Tal aprobación debe obtenerse ***antes*** del comienzo del estudio[13,25,28].

En la práctica, la labor de estos comités consiste en representar a la sociedad, para garantizar que los principios de la Declaración de Helsinki y de otras declaraciones bioéticas pertinentes se apliquen a todas las investigaciones en las que participan sujetos humanos, así como velar por que todos los medicamentos, las vacunas y los equipos investigados cumplan con adecuados estándares de seguridad. Por otra parte, es también tarea de los comités vigilar que la investigación se efectúe siguiendo los protocolos aprobados. Cualquier desviación sustancial del protocolo aprobado debe ser evaluada y aceptada para su ejecución por el mismo comité que estudió la propuesta original. En el mismo sentido, el comité está autorizado para solicitar a los investigadores presentaciones sobre la marcha de los estudios autorizados, así como sus resultados parciales o definitivos.

La constitución, la organización y los métodos seguidos por cada comité dependen, fundamentalmente, de las políticas de la organización de la práctica médica y de la investigación de la institución, así como del grado de autonomía acordado con los investigadores médicos. En general, se recomienda que los comités incluyan a médicos y a científicos, aunque también, a

representantes de otras profesiones, tales como enfermeras, abogados, sacerdotes, expertos en ética y representantes de la comunidad.

Es muy importante tomar en cuenta que la evaluación ética de un proyecto incluye tres aspectos fundamentales:

1. **Los aspectos éticos propiamente dichos,** que pueden evaluarse con el apoyo de las declaraciones internacionales, regionales y locales, y con la discusión de los aspectos no regulatorios en el comité. Como ejemplos, la decisión de incluir o no a niños en el estudio tiene claros referentes en los documentos básicos, y ante la duda, por una condición especial, se hará la consideración en el comité, teniendo en cuenta las opiniones calificadas de los representantes de las distintas partes (profesionales científicos, bioeticistas, representantes de los pacientes y de la comunidad).
2. **Los aspectos científicos. Es tarea del comité** evaluar si son o no apropiados los planteamientos de la justificación y la pregunta de investigación, así como las decisiones sobre las intervenciones que se llevarán a cabo. Un comité debe tener certeza de que hay justificación científica para el estudio, y de que las intervenciones son adecuadas y proporcionadas. Un comité conformado tan solo por bioeticistas no podría evaluar este aspecto, y por tal motivo, la composición debe ser amplia y comprehensiva.
3. **Los aspectos metodológicos. Debe considerarse** que hacer un estudio con un tamaño de muestra insuficiente es claramente antiético, pues se gastan recursos, se expone a los participantes a riesgos o incomodidades y no hay garantía de validez en los resultados. Y si el tamaño de muestra es excesivo, por un cálculo inapropiado, se desperdiciarán recursos y tiempo, y se expondrá a más sujetos de los necesarios… Igualmente, un diseño inapropiado, o bien, un control incorrecto de sesgos o de factores de confusión, hará que los resultados no puedan ser válidos, y esto es una falla ética muy significativa.



Los comités deben, entonces, evaluar los aspectos éticos, científicos y metodológicos de cada proyecto. Y esas evaluaciones no pueden conducirse de manera separada.

Un estudio que tenga defectos metodológicos serios no es ético, como no lo es que tenga defectos científicos, por cuanto expone a los participantes al riesgo de una investigación que arrojará resultados que no van a ser válidos, o va a exponer a los participantes a riesgos innecesarios, o a consumir recursos de manera incorrecta. Normalmente, los comités institucionales consideran (y así debe ser) de forma simultánea los aspectos éticos, metodológicos y científicos de cada proyecto.

Vale la pena resaltar que el comité debe evaluar simultáneamente todos los aspectos, para lo cual debe haber en su constitución personas que velen por la pulcritud en el área académica, en la cual deben garantizarse la pertinencia y la relevancia de la pregunta, así como en el área metodológica, y, por supuesto, en el área ética propiamente dicha.

El estudio debe generar un conocimiento útil y generalizable, y no debe repetir innecesariamente experiencias previas. También deben garantizarse la validez interna y externa del proyecto: que sean correctos todos los aspectos, como la formulación de la hipótesis, la selección de variables de desenlace, el control de las variables de confusión, el análisis estadístico y muchos otros aspectos, pues son claros determinantes de la validez de los resultados de un estudio. Adicionalmente, y no menos importante, el uso de recursos, el respeto por el medio ambiente, los gastos y los dineros involucrados y los aspectos logísticos deben ser cuidadosamente controlados.

Por lo anterior, y para determinar la validez externa, o generalizabilidad, el comité debe tener expertos en el área metodológica. Por último, es necesario que el comité pueda evaluar, simultáneamente, los aspectos bioéticos de la investigación; no solo los obviamente relacionados con la ética (CI, evaluación de riesgos, confidencialidad, respeto), sino también los relacionados con las áreas académicas y metodológicas que tengan claro tinte bioético: un estudio con defectos metodológicos, como ya se comentó —con un tamaño de muestra desproporcionado, con una medición inapropiada de variables de desenlace o con un análisis bioestadístico inadecuado—, es tan poco ético como un estudio sin CI, sin justificación científica o con exposición a riesgos desmedidos a los participantes.

Los aspectos académicos, metodológicos y bioéticos deben ser adecuadamente evalua-

dos en todas las etapas del estudio: tanto en la preparación y en la aprobación final como en la conducción, la conclusión y la utilización de los resultados, que incluye los aspectos éticos de las presentaciones en público, la difusión y la autoría en las publicaciones.

En relación con el documento para firmas del CI, **este es un documento difícil de elaborar;** especialmente, para investigadores sin experiencia, de tal modo que el comité puede colaborar con los investigadores en la elaboración de los formularios de consentimiento desarrollando un trabajo en equipo productivo, que proteja a los sujetos, sin obstaculizar la investigación.



El comité debe preparar modelos de documentos de CI que sirvan de ejemplo, y distribuirlos entre los investigadores y hacer énfasis en los requerimientos que la norma local exige a los documentos del CI.

Asimismo, debe dejar clara la diferencia entre el lenguaje sugerido y el lenguaje requerido en el documento y definir a qué lenguaje se atiene el comité cuando revisa el documento. Sí el comité tiene ciertos parámetros (***lenguaje requerido***), **el investigador debe conocerlos de** antemano. El comité debe dejar claro que el lenguaje alternativo es aceptable, pero existen unos mínimos que deben estar incluidos en el documento final.

El comité debe desarrollar ejemplos que expliquen cómo se deben manejar ciertos temas: por ejemplo, si el documento del CI considera el uso de muestras biológicas, se deben exponer ejemplos y qué se espera del documento para esa circunstancia particular. El comité también debe indicar al investigador que el documento de CI debe incluir los riesgos, los beneficios y los procedimientos explicados sin contradicciones, y siempre referenciados al protocolo de investigación, y proveer al investigador una lista de información mínima que debe estar presente en el documento **(tabla 36.1)**, información que debe ajustarse a las diversas actividades del protocolo. Por ejemplo, explicación sobre el uso del placebo, sobre la aleatorización, permiso para grabación en video, movilización de muestras genéticas, pruebas de embarazo, estudios colaborativos con otros investigadores, asentimientos, etc. Es decir, el comité debe "adelantarse" y colaborar en el proceso, y no ser, simplemente, un receptor y un evaluador pasivo.

El comité puede preparar documentos de CI para distintos escenarios: investigación clínica, investigación en colegios, comunidades, etc., y desarrollar documentos y parámetros para diferentes tipos de investigación, como experimentos clínicos, entrevistas, grupos focales, etnografías, etc., en colaboración con los investigadores.

Dado que una vez evaluado y aprobado el protocolo de investigación se iniciará una intensa actividad, que enfrenta diversas dificultades logísticas y de ejecución, el seguimiento del estudio es un aspecto imperativo para el comité, que debe encargar a uno de sus miembros el seguimiento del estudio y la comunicación constante con el investigador, del tal modo que se verifique el cumplimiento de lo planteado en el sometimiento del estudio y se garantice la protección del sujeto durante el desarrollo de la investigación. Como ya se mencionó, dicho seguimiento incluye verificar la difusión de los resultados, independientemente de si estos favorecen o no las hipótesis de estudio.

Referencias

1. Shuster E. Fifty years later: the significance of the Nuremberg Code. N Engl J Med. 1997;337:1436-40.
2. WMA Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects [internet]. 2013 [citado 2014 jul. 7]. Disponible en: http://www.wma.net/en/30publications/10policies/b3/
3. Pacto Internacional de Derechos Civiles y Políticos [Internet]. [citado 2014 jul. 7]. Disponible en: http://www2.ohchr.org/spanish/law/ccpr.htm
4. ASH. The Belmont Report [internet]. [citado 2014 jul. 7]. Disponible en: http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html
5. CIOMS. International Ethical Guidelines for Biomedical Research Involving Human Subjects 1993 [internet]. [citado 2014 jul. 7]. Disponible en: http://www.cioms.ch/index.php/printable-publications?task=view&id=37&catid=57.
6. ICH. Good Clinical Practice 1996 [internet]. [citado 2014 jul. 7]. Disponible en: http://www.ich.

org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html.
7. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000;283:2701-11.
8. Emanuel EJ, Wendler D, Grady C. An ethical framework for biomedical research. Oxf Textb Clin Res Ethics. 2008;123-35.
9. Marsh V, Kamuya D, Rowa Y, et al. Beginning community engagement at a busy biomedical research programme: Experiences from the KEMRI CGMRC-Wellcome Trust Research Programme, Kilifi, Kenya. Soc Sci Med. 2008;67:721-33.
10. Allman D, Ditmore MH, Kaplan K. Improving ethical and participatory practice for marginalized populations in biomedical HIV prevention trials: lessons from Thailand. Plos One. 2014;9:e100058.
11. Coughlin SS, Beauchamp TL. Ethics, scientific validity, and the design of epidemiologic studies. Epidemiology. 1992;3:343-7.
12. Edwards HA, Hifnawy T, Silverman H. Enhancing research ethics review systems in Egypt: The focus of an international training program informed by an ecological developmental approach to enhancing research ethics capacity. Dev World Bioeth. 2014.
13. CIOMS. International guidelines for ethical review of epidemiological studies. Ginebra: CIOMS; 1993.
14. Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects. Bull Med Ethics. 2002;182:17-23.
15. Colombia, Ministerio de Salud. Resolución 8430 de 1993, por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Bogotá: Minsalud; 1993.
16. Hill AB. The continuing unethical use of placebo controls. N Engl J Med. 1994;331:394-8.
17. Congreso de Colombia. Ley 23 de 1981, por la cual se dictan normas en materia de ética médica. Bogotá: 1981.
18. Colombia Ministerio de la Protección Social. Resolución 2378 de 2008, por la cual se adoptan las Buenas Prácticas Clínicas para las instituciones que conducen investigación con medicamentos en seres humanos. Bogotá: Minprotección; 2008.
19. INVIMA. Buenas Prácticas Clínicas [internet]. 2014 [citado 2014 ago. 9]. Disponible en: https://http://www.invima.gov.co/index.php?option=com_content&view=article&id=991&Itemid=326.
20. Colombia, Instituto Nacional de Medicina Legal y Ciencias Forenses. Resolución 485 de 2002, por medio de la cual se reglamenta el procedimiento para la entrega de cadáveres y componentes anatómicos que se obtengan de los mismos, para fines de docencia e investigación. Bogotá: 2002.
21. Colombia, Ministerio de Salud. Resolución 1995 de 1999, por la cual se establecen normas para el manejo de la Historia Clínica. Bogotá: Minsalud; 1999.
22. Congreso de Colombia. Ley estatutaria 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. Bogotá: 2012.
23. De Nuremberg C. Tribunal Internacional de Núremberg, 1947. Trials War Crim Nuremberg Mil Trib Control Counc Law. 1949;10:181-2.
24. Hellman S, Hellman DS. Of mice but not men: problems of the randomized clinical trial. N Engl J Med. 1991;324:1585-9.
25. Schafer A. The ethics of the randomized clinical trial. N Engl J Med. 1982;307:719-24.
26. Modi N, Vohra J, Preston J, et al. Guidance on clinical research involving infants, children and young people: an update for researchers and research ethics committees. Arch Dis Child. 2014;99:887-91.
27. Culliton BJ. Clinical investigation: an endangered science. Nat Med. 1995;1:281.
28. Conroy S, Choonara I, Impicciatore P, et al. Survey of unlicensed and off label drug use in paediatric wards in European countries. BMJ. 2000;320:79-82.
29. Angell M. The ethics of clinical research in the Third World. N Engl J Med. 1997;337:847-9.
30. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;317:141-5.
31. Wald N. Ethical issues in randomised prevention trials. BMJ. 1993;306:563-5.
32. Passamani E. Clinical trials: Are they ethical? N Engl J Med. 1991;324:1589-92.
33. Lucassen A, Houlston RS. The challenges of genome analysis in the health care setting. Genes. 2014;5:576-85.
34. Peterson-Iyer K. Pharmacogenomics, ethics, and public policy. Kennedy Inst Ethics J. 2008;18:35-56.
35. Sciences C for IO of M, others. The Declaration of Inuyama. Bull Med Ethics. 1991;8.
36. Abbing HDCR, UNESCO. International Declaration on Human Genetic Data. Eur J Heal Law. 2004;11:93-107.
37. UNESCO. Universal Declaration on the Human Genome and Human Rights. Nurs Ethics. 2001;8:276-82.

CAPÍTULO

37 Autoría

ÁLVARO J. RUIZ MORALES

INTRODUCCIÓN

Puede hablarse de ***autoría justificada*** cuando cada autor ha hecho contribuciones significativas al diseño del estudio, a su ejecución, al análisis de los datos y al desarrollo del manuscrito. Aunque parece simple y claro, hay muchas dudas alrededor del tema y muchas desviaciones. Se duda, por ejemplo, de si puede justificarse la autoría por participar solamente en el reclutamiento de individuos o en la recolección de la información.

En general, la autoría significa una declaración pública y formal de responsabilidad por el trabajo realizado y publicado, y de que se garantiza que los datos han sido representados de manera correcta y honesta. Y de que se asumirá la responsabilidad sobre cualquier parte del trabajo, por todos los autores.

La investigación está basada en la confianza. Tanto en su ejecución como en la presentación de sus resultados, sea de forma oral o escrita, la comunidad científica cree en la buena fe y en la veracidad de lo informado.

Se asume que los investigadores han sido honestos y cuidadosos en la ejecución del proyecto, y que los resultados se trataron, se analizaron y se informan de manera transparente, no sesgada y completa[1]. Se asume que se informa la verdad, solo la verdad y nada más que la verdad, y que cualquier aspecto relevante adicional, como potenciales conflictos de interés, relaciones profesionales o comerciales pertinentes o información importante sobre individuos que se han retirado del estudio, por ejemplo, se informan clara y oportunamente.

El acelerado desarrollo de la medicina y de la investigación ha traído como consecuencia el aumento exponencial de publicaciones que informan investigación. Y las necesidades creadas por la academia, por el ambiente laboral, por las universidades y las asociaciones científicas y por el medio profesional han reforzado la necesidad de tener publicaciones, y cuantas más haya, y en medios más prestigiosos, tanto mejor.

Y se ha producido, por otra parte, un cambio en las publicaciones científicas, por la complejidad de los proyectos, por la interacción entre grupos de investigación y por la creciente interdisciplinariedad: de la tradición de un único autor, que aún persiste en algunas áreas (particularmente, las humanidades), se pasó a la tradición de múltiples autores.

Se hace cada vez más necesario que el mismo enfoque de transparencia y honestidad que se espera de la conducción de la investigación se aplique a las decisiones sobre autoría, para que se cumpla con el lector y con la comunidad

científica al ofrecer información que pueda ser aceptada con confianza. Con la certeza de que allí está la verdad, completa y exclusiva, con la información suficiente para una adecuada interpretación.

No solo hay cada vez más revistas científicas y más artículos, con cifras en acelerado crecimiento, sino que cada vez hay más autores por artículo. Y se hace indispensable que haya claridad en las reglas para la definición de quién *puede* ser autor de un artículo, quién *debe* ser autor, cuál debe ser el orden correcto de los autores y cuáles son los aspectos relevantes para la definición de autoría.

PROBLEMAS CON LA AUTORÍA

Son múltiples los problemas descritos con la autoría que ameritan respuestas claras y justas a preguntas como: ¿quién puede ser autor de una publicación?, ¿quién puede figurar?, ¿quién debe aparecer y a quién se le deben dar agradecimientos o reconocimientos?, ¿quién puede publicar?, ¿qué responsabilidades tienen los autores, y cuáles, el autor principal? y ¿qué se puede publicar?

Los criterios precisos y apropiados para la solución de esas preguntas permitirán atribuir, en justicia, la propiedad intelectual a quien le corresponde, harán claridad en las reglas de promoción académica y facilitarán las búsquedas a las personas que utilizan la literatura científica y a quienes hacen investigación secundaria. Por ello, la publicación debe estar regida por la veracidad, la justicia, la búsqueda de beneficio y la ausencia de daño[2].

Recientemente se ha descrito una enfermedad: la *poliautoritis giftosa*[3], cuya etiología es el trastorno que ha generado el aumento incontrolado del número de autores y de los vicios que se cometen con frecuencia en el establecimiento de la autoría. Sus manifestaciones son un número elevado de autores y la muy desigual contribución de cada uno de ellos (en muchos casos, inexistente).

Esta proliferación de autores en los artículos ha llevado a extremos como tener 25 autores en un experimento clínico simple o 16 autores en el reporte de un caso clínico. Las consecuencias de la práctica de autoría múltiple e inmerecida son muchas y con implicaciones muy serias en aspectos de justicia laboral, de reconocimiento académico, de distribución de recursos y de respeto a las normas éticas y de propiedad intelectual.

ORÍGENES DE LOS PROBLEMAS EN AUTORÍA

Aumento de la actividad investigativa y de la colaboración

El crecimiento acelerado en la investigación, el desarrollo de medicamentos y de intervenciones terapéuticas y la rápida aparición de estrategias diagnósticas avanzadas y de tecnologías complejas han sido los determinantes principales del número creciente de publicaciones científicas.

Los avances en informática y en comunicaciones han facilitado, por su parte, la amplia y rápida difusión de la información y la disponibilidad, cada vez mayor y más fácil, de las revistas y los artículos científicos.

Por decisiones logísticas y de racionalización del uso de recursos, así como por el aumento en las redes de investigadores, la investigación ha ido evolucionando, para convertirse, cada vez más, en una actividad colectiva más que individual. Esto permite una mejor eficiencia en el aprovechamiento de recursos, más facilidad para el reclutamiento de individuos —lo que resuelve problemas de tamaño de muestra— y mayor diversidad en la representatividad de la población de estudio lo que redunda en una mejor validez externa, entre muchas otras ventajas logísticas y de intercambio académico.

Gracias a esta tendencia, pueden evitarse los estudios de bajo poder estadístico o con escasa variabilidad, se facilitan los análisis estratificados, ya que se alcanzan números significativos en los diferentes grupos de interés, aumenta la colaboración regional e internacional y se ahorran recursos al permitir la unión de esfuerzos y evitar duplicaciones o repeticiones innecesarias.

Difusión de criterios de evaluación académica y laboral

Las universidades, los hospitales y las sociedades científicas quieren estimular la producción intelectual, y, a la vez, quieren utilizar esta ac-

tividad como elemento de juicio y calificación de sus miembros. Los estatutos docentes y los criterios para contratación y promoción en las entidades científicas y hospitalarias se han venido perfeccionando, y los requisitos son cada vez más claros y exigentes.

Como resultado, ha habido un aumento en el número y en la calidad de las publicaciones, el perfil de médicos y los profesores ha evolucionado hacia un mejor entrenamiento en metodología y en bioestadística, hay una mayor actividad investigativa, más y mejores proyectos presentados en congresos científicos y más artículos publicados en un número siempre creciente de revistas científicas.

Además de lo anterior, tanto en las universidades como en las instituciones de salud y en las sociedades científicas se exige con mayor frecuencia que haya educación continuada, eficiencia demostrada en el ejercicio de la profesión y mayor cantidad y calidad de investigación publicada. La producción de investigación requiere difusión apropiada, ágil y oportuna; y este es el principal estímulo para la publicación, y se constituye en su objetivo principal.

Pero hay otras razones válidas y poderosas para publicar: se publica para darse a conocer, para dar visibilidad al grupo de investigadores o a la institución, para presionar la incorporación del conocimiento en la práctica diaria o en las políticas y en la toma de decisiones. También, para estimular la producción de más investigación y el desarrollo de nuevas líneas, con el fin de facilitar y promover el contacto con otros investigadores.

Y hay un beneficio adicional, válido y legítimo: el componente de prestigio y reconocimiento académico (y, en ocasiones, financiero) que puede lograrse a través de las publicaciones, tanto en la comunidad científica internacional como en el ámbito local.

Para las personas vinculadas con el área académica, las publicaciones agregan elementos evaluables a la producción intelectual, que pueden traducirse en puntos para la promoción académica, para la obtención de estímulos, e, incluso, para la retribución económica. En el currículo del profesional, las publicaciones, como aspecto evaluable, permiten jerarquizar los candidatos para posiciones de trabajo o para promociones, decisiones sobre retribuciones o bonificaciones.

Consecuencias del uso incorrecto de los criterios de autoría

Las consecuencias de la autoría inapropiada se derivan directamente de los beneficios. Un profesional a quien se le regale autoría va a ganar prestigio, renombre, posición académica, puntos en su evaluación laboral y académica, estímulos económicos y otros beneficios totalmente injustificados. Y no solo va a tener una retribución injusta, sino que en muchos casos tendrá elementos para la competencia desleal con sus colegas, y recibirá dinero y beneficios que no merece, y que dejarán de percibir otras personas que sí los merecen. La autoría inapropiada proporciona beneficios no justificados y éticamente discutibles a algunos, y quita beneficios y pone en situación de desventaja a otros.

Puede ocurrir, también, lo contrario: que se le niegue autoría a quien la merece, o que se le relegue en el orden de autores a una posición que no corresponda a sus aportes. O que se oculte información relevante sobre potenciales conflictos de interés o relaciones con la industria farmacéutica, o con posesión de patentes o de intereses comerciales.

Problemas con la autoría

El principal problema, origen de los demás, estriba en el poco conocimiento de los criterios propuestos de definición de autoría, en la poca difusión que se les da y en el poco celo en la vigilancia de su cumplimiento, que se presentan con frecuencia en instituciones académicas y científicas y en el ámbito laboral.

Además de lo anterior, y como resultado del número mayor de proyectos en los que participa un número también cada vez mayor de colaboradores, se presentan dificultades en la definición de las responsabilidades y los méritos que corresponden a las diferentes personas que estuvieron involucradas en el proyecto[1]. Y no es raro que haya imprecisiones en las decisiones sobre autoría, como las mencionadas en la **tabla 37.1**.

Inclusión injustificada

La autoría, que es un reconocimiento del trabajo realizado, tiene además beneficios de los cuales será injusto disfrutar si no se cumplen los criterios para merecerla. Quien cumpla

Tabla 37.1 Problemas en la autoría
Inclusión injustificada de una o más personas
Falta de una o de varias personas que merecen autoría
Orden de los autores que no corresponde con la magnitud de su colaboración
Omisión de la filiación o de información en uno o en varios autores
Publicaciones repetidas
Autocitación o citas injustificadas o inapropiadas

Fuente: Elaboración del autor.

los criterios debe ser autor, y solamente quien los cumpla puede aparecer como tal, puesto que la autoría debe ser no solo algo merecido y ganado, sino explícito. Hay justificación para la autoría cuando se ha hecho una contribución sustancial, de acuerdo con los criterios que se han propuesto y difundido, a una publicación. La autoría se define por el trabajo realizado, no por la posición ocupada[4].

Hay varias modalidades de inclusión injustificada:

a. **Autoría impuesta:** El director del servicio, del departamento o del centro obliga a los miembros de su grupo a incluirlo en las publicaciones que haga cualquiera de ellos, aunque su único mérito consista en el ser el jefe o en estar a cargo de la administración. No solo es una autoría injusta e inmerecida, sino que se trata de una clara imposición violenta y atrevida. Una opción bastante peor, sin embargo, es la imposición de incluir a todos los miembros del servicio o de la unidad académica en todas las publicaciones. Es una mezcla de imposición y de la modalidad siguiente (autoría como regalo), que tiene los poco honestos propósitos de aumentar el número de publicaciones del grupo y de cada uno como individuo. Es claro que trabajar en un área clínica no hace al especialista dueño de los pacientes, y que se haga un trabajo de investigación con individuos de esa área no justifica de ninguna manera la autoría colectiva. Tampoco la justifica el que el grupo comparta la responsabilidad del cuidado médico de los pacientes, pues la investigación tiene propósitos diferentes y debe estar claramente diferenciada de la atención habitual en salud.
b. **Autoría donada:** Invitar a colegas, a amigos, a miembros del servicio o de la misma unidad académica a ser autores como cortesía no solo incumple los criterios de justificación y de equidad, sino que reparte, inapropiadamente, los recursos, los beneficios y el honor que se derivan de aparecer en la lista. Con frecuencia se incluye a una o más personas que no participaron, pero a quienes se les debe algún reconocimiento. En estos casos, bien porque esas personas apoyaron la consecución del financiamiento, o bien, porque sustituyeron a los autores durante alguna responsabilidad clínica, o porque, de alguna forma tangencial, ayudaron al proyecto, lo apropiado es mencionar su aporte como un agradecimiento o un reconocimiento. Ofrecer autoría inmerecida para que el recipiente, a su vez, incluya al oferente en sus proyectos no deja de ser una asociación injusta y claramente inapropiada.
c. **Autoría como retribución:** Mencionada en la modalidad anterior, es una sobreexpresión de gratitud o de reconocimiento. No hay ninguna justificación para ofrecer autoría a quien no la merece, y si hay razones para agradecer a quien hizo aportes insuficientes para que sea autor, la sección de agradecimientos cumple ese propósito.
d. **Autoría por compromiso:** Por tradición del servicio o del departamento, y de manera incorrecta e injusta, se hacen pactos, extremadamente objetables desde el punto de vista ético, de incluir como autores a todos los miembros de un servicio, de un departamento o de una unidad. "Todos" los miembros de la unidad académica o de trabajo figuran en los trabajos de todos, con el fin de aumentar el número individual de publicaciones, para corresponderse mutuamente favores, para tener figuración local, nacional o internacional aumentada, para mejorar la figuración del grupo o de la unidad académica, etc. No tiene ningún sentido, ni debe ser una justificación, el que algunos protegen el tiempo de otros para sus proyectos, y que se espera que ello se retorne en reciprocidad.
e. **Autoría honoraria:** Con alguna frecuencia el director del servicio, o una persona de gran prestigio en el área, recibe una invitación, aunque no haya contribuido al proyecto. Comúnmente, se asocia ello a una autoría fantasma, la cual se explicará más adelante: se evita a propósito el mencionar a quienes han hecho el estudio y se invita a profesio-

nales reconocidos para que aparezcan como autores. Incluso, en algunos casos se incluye a algunas personas sin su consentimiento.

f. **Autoría con interés:** Un profesional de prestigio, o con gran reconocimiento en el área, o con un amplio número de publicaciones, puede recibir una invitación para ser autor con el objeto de facilitar la recepción para que se publique el artículo, o para aumentar las probabilidades de llegar a una revista. O se invita como autor a quien hace un procedimiento (microscopía electrónica, polisomnografía, potenciales evocados) buscando que apoye la realización de exámenes para otro estudio, o para que ofrezca tarifas especiales.

g. **Autoría por realización de pruebas o de estudios:** De manera equivocada, se piensa que haber contribuido a la realización de exámenes paraclínicos justifica la autoría. Quien hizo 230 ecocardiogramas para un estudio, pero no hizo ninguna otra contribución al estudio, no cumple criterios para ser autor. Según esta teoría, los bacteriólogos y los radiólogos deberían ser siempre incluidos como autores de todos los estudios en los que haya pruebas paraclínicas, al igual que los electrofisiólogos, los neumólogos, los gastroenterólogos, etc. En algunos casos, hay razones suficientes para un agradecimiento explícito. Pero este criterio único no justifica la inclusión como autor.

Exclusión injustificada

Se ha descrito como ***autores fantasma*** a quienes hacen contribuciones intelectuales sustanciales y suficientes, pero no aparecen listados. Hay, al menos, dos razones para que se los omita:

a. **Conflictos de interés:** Habitualmente, la presencia de un miembro de una compañía farmacéutica o de una empresa comercial que produce equipos o insumos médicos va a ser vista como un evidente conflicto de interés. Para evitar esta apreciación, y de manera inapropiada, se oculta la participación de personas relacionadas con las compañías patrocinadoras, o se oculta el apoyo económico o científico de una empresa. En este caso, no solo se niega autoría a quien la merece, sino que se engaña a los lectores y se les impide conocer potenciales conflictos de interés.

b. **Autores contratados:** En ocasiones, se contrata a alguien para que lleve a cabo el estudio, incluidos los análisis y la escritura del manuscrito, pero se les deja claro que no figurarán como autores, ya que se ofrecerá la autoría a profesionales de renombre.

Orden inapropiado de los autores

Como ya se comentó, cada vez es más frecuente que haya estudios producidos por equipos de investigación, de manera que la coautoría es algo cada vez más usual. Más aún, el número de autores ha aumentado con el paso del tiempo. Por esas razones, debe buscarse que el orden responda a criterios sólidos, y los cuales, a su vez, se basen en la cuantificación del trabajo realizado, de la responsabilidad cumplida o de la importancia de los aportes.

Dado que el primer autor será generalmente el citado, esta es una posición en el orden de autoría que genera prestigio, y que se busca siempre. Sin embargo, su importancia también lleva a que se la ofrezca con criterios diferentes de los realmente justos.

Aunque parezca obvio, el autor principal debe ser quien hizo las mayores contribuciones al estudio, y el orden siguiente debe expresar la magnitud decreciente de contribuciones. No debe haber ningún criterio externo al proyecto de investigación, a su conducción, su análisis ni su escritura que deba cumplir un papel en la selección del primer autor. Ni debe buscarse mejorar las posibilidades de publicación, ni obedecer a las insinuaciones de algún jefe o de un director deshonesto, ni hacer tráfico de influencias o favores, ni pagar o agradecer tiempo protegido, o financiación o la ayuda logística, para lograr algún equipo o el reclutamiento de pacientes.

También en el orden puede faltarse a la verdad o a la justicia, y, por diversas razones, este tema es fuente de insatisfacción y de conflictos, o que no se evalúa con suficiente atención. Puede haber diversas manifestaciones de dicho defecto:

1. **El primer autor no necesariamente es quien** hizo las mayores contribuciones, sino la persona de mayor rango jerárquico o de mayor prestigio. Es evidente que la magnitud de la contribución debe ser el criterio único.
2. **El último autor, a menudo, es quien merecería** ser el primero. Frecuentemente se relega a quien ha hecho los mayores aportes al último puesto, el cual, para algunos, tiene un significado de alto prestigio. Y, además, se ha

establecido un uso indebido del último puesto en el listado, para dejar allí al jefe del grupo de investigación, al director del servicio o del laboratorio o a la persona de mayor prestigio, independientemente de su contribución. Esta práctica tiene tanta aceptación que algunas revistas, incluso, citan a un número fijo de autores (por ejemplo, cinco) y al último, aunque haya doce autores en total. En algunos casos, se trata de un equipo que tiene una línea definida de investigación que dirige una cabeza científica, quien, además, tiene el prestigio, la facilidad para conseguir financiación y a quien se le abren las puertas de las revistas para publicación. Cuán justificada es dicha conducta es un tema de difícil abordaje. Se cumpliría más justamente con el respeto a los derechos de todos si el orden fuese estricto y en orden decreciente. Debe anotarse que esta es una práctica difundida y aparentemente aceptada en algunos países (particularmente, en Europa).

3. Aunque una persona merezca el puesto como primer autor, si es un estudiante de pregrado o de posgrado, se la relega a posiciones secundarias. Debe recordarse que no son la posición jerárquica ni los títulos académicos los que determinan el orden. Debe serlo, estrictamente, el mérito en la contribución al estudio y al manuscrito.
4. *Orden variable según la conveniencia:* A menudo, el primer autor es una persona para una circunstancia (presentación del trabajo en un congreso); otra para la publicación en una revista, y una más, para la publicación en otra revista. No tiene sentido que haya variación en el orden, por cuanto es la contribución de los autores la que debe definir este aspecto. Esta modalidad viola varios principios éticos y de derechos de autor. Hay estudios en los que la solicitud de financiación trae como primer autor a un investigador con prestigio reconocido y antecedentes de trabajos con la agencia financiadora, lo cual facilita la decisión de brindar apoyo económico. Sin embargo, ese mismo proyecto de financiación se presenta con otro autor en primer lugar ante las autoridades de una universidad, para ganar puntos en el escalafón docente. En el congreso de la especialidad, el primer autor es el más joven del grupo, para que el estudio sea aceptado en el concurso que hace el congreso de jóvenes investigadores. Y en las presentaciones posteriores en reuniones científicas, el orden cambia a conveniencia de los presentadores y según quién esté llevando la voz cantante. Cualquiera de estas variaciones en el orden debe ser considerada una falta ética, y esta práctica inadmisible debe ser denunciada y abolida.

Omisión de información relevante para el análisis de conflictos de interés

Bien por buscar apariencia de neutralidad o por ocultar conflictos evidentes de interés, se omite a uno o más autores, o se oculta su filiación. Las dos costumbres quitan transparencia al escrito e impiden que el lector tenga los elementos de juicio para evaluar la información.

Debe recordarse que la participación en un estudio (generalmente, un experimento clínico) de la industria farmacéutica o de una empresa que produzca equipos o dispositivos puede producir presiones, las cuales pueden afectar la validez de los resultados o, al menos, la credibilidad de ellos. Abbenheim *et al.* publicaron un estudio favorable a un nuevo medicamento para la obesidad: la dexfenfluramina[5], producida por Interneuron Pharmaceutical Inc. A pesar de que los dos autores del artículo trabajan para la empresa que se encargaba de la comercialización y la producción del medicamento en Estados Unidos y Europa, no hicieron explícito este claro conflicto de intereses. Por efecto de la publicación, las acciones de la compañía aumentaron el 13 %. Es obvio que se violó un principio básico al no mostrar explícitamente la relación, pues así se quita al lector la posibilidad de evaluar la posible presencia de sesgos[6].

Sin embargo, la mencionada participación no significa automáticamente que haya conflictos de interés[7]. La información debe ser consignada de manera completa y correcta, para que el lector pueda juzgar si puede o no haber un conflicto, o si puede haber influencia inapropiada o manipulación de la información[8]. Esta evaluación se hace mediante el análisis cuidadoso y crítico del artículo, y debe llevarse a cabo en todos los casos.

Con frecuencia se considera que la participación de la industria farmacéutica implica la presencia de sesgos, o que hay un riesgo mucho más alto de manipular la información. Debe considerarse que en todos los casos debe hacerse la misma evaluación, crítica y a fondo, ya que un

estudio iniciado, conducido y financiado por un investigador independiente o por una institución académica puede también tener sesgos, o intereses suficientes en aspectos no académicos que sean verdaderos conflictos de interés.

Publicaciones repetidas

Se dice que la publicación de un artículo es redundante o repetida cuando hay coincidencia sustancial con otro ya publicado. El artículo nuevo debe explorar otras posibilidades, pues, justamente, eso espera y merece el lector, a menos que se especifique claramente que se trata de una repetición justificada y autorizada tanto por los autores como por el editor de la revista en la que se hizo la primera publicación[9].

En general, no deben considerarse publicaciones repetidas las que siguen a presentaciones preliminares en forma de resumen o de afiche para un congreso, o a la presentación en una reunión científica o a la publicación de reportes de prensa. Se consideran publicaciones preliminares los reportes con detalles, tablas y figuras que se envían a medios masivos de comunicación o a agencias gubernamentales. Cuando sea necesario hacerlo, por implicaciones éticas o de salud para la comunidad, habrá siempre un acuerdo al respecto entre el editor y los autores del estudio. Como ejemplo, un hallazgo de gran impacto en la salud puede ser publicado de manera preliminar, para buscar su difusión y su implementación inmediatas, sin que se deba esperar a la salida de la publicación, que puede tomar un tiempo largo (en ocasiones, más de un año). En este caso, un informe preliminar con datos que soporten la recomendación puede ser publicado o enviado a medios de comunicación o de difusión, pero siempre debe haber acuerdo entre los autores y los editores de la revista que tendrá la publicación completa.

El autor se obliga a declarar toda publicación previa existente del mismo trabajo o de uno similar. Son aceptables las publicaciones secundarias o repetidas (particularmente, en distintos idiomas o en diferentes países), siempre y cuando se considere que tienen justificación y beneficio, y si se cumplen las siguientes condiciones[10]:

1. **Los autores han recibido aprobación de los** editores de las dos revistas (el editor de la segunda revista debe tener acceso a la versión primaria).
2. **Se respeta la prioridad de la publicación** primaria en un intervalo en el que los dos editores se ponen de acuerdo con los autores.
3. **El artículo para la publicación secundaria tiene** como objetivo un grupo diferente de lectores; a menudo, podría bastar una versión abreviada.
4. **La versión secundaria refleja fielmente los** datos y las interpretaciones de la versión primaria.
5. **La versión secundaria informa a los lectores,** a los pares y a las agencias de documentación que el artículo ha sido publicado en su totalidad o en parte en otra fuente, por ejemplo, con una nota que puede decir: "Este artículo se basa en un estudio reportado inicialmente en la revista [título de la revista, con la referencia completa]", y la versión secundaria debe tener la referencia primaria.
6. **El título de la publicación secundaria debe** indicar que es una publicación secundaria (completa, abreviada, o traducida) de una primaria.

Como nota de interés, la National Library of Medicine (NLM, o Biblioteca Nacional de Medicina) de los Estados Unidos (la misma que produce MedLine) no considera las traducciones como otra publicación, y no las cita ni las indexa cuando el artículo original fue publicado en una revista también indexada en MedLine[10]. Si un artículo se publica en múltiples idiomas en la misma revista, MedLine lo cita agregando al final de la cita una nota que dice, por ejemplo: "Artículo en inglés, francés y holandés".

Modalidades inadecuadas de autoría consisten en la publicación innecesaria del mismo estudio o artículo varias veces, sin justificación plena; el fraccionamiento inadecuado de un estudio para obtener múltiples publicaciones, o la variación sin respaldo del orden de autoría en varias publicaciones, para dar oportunidad a varios de aparecer como primeros autores del mismo estudio[11]. **Los casos más frecuentes de** autoría inadecuada son:

- Publicaciones múltiples, en diferentes revistas, del mismo artículo.
- Publicación innecesaria del mismo artículo en diferentes idiomas.
- Artículos parciales sobre el mismo tema, y en los cuales no se justifica la fragmentación, con el objetivo de tener más publicaciones.

- Repetición de un artículo en la misma revista, con intervalos variables. La razón de este fenómeno tiene que ver con las causas de los otros problemas mencionados: la necesidad urgente de tener un número elevado de publicaciones con el fin de ganar nombre, el afán de puntaje en algún escalafón, tener o mejorar las posibilidades de ascensos, o mantener el trabajo o la financiación. A menos que haya una justificación clara, diferentes autores (Kassirer, Huth, Huston) consideran que dicha práctica es inadecuada y falta a la ética[8].

Sin embargo, hay justificaciones para el fenómeno de publicaciones múltiples que deben tenerse en cuenta, pues al proceder así no en todos los casos se falta a las leyes de autoría:

- La publicación del mismo estudio en dos revistas dirigidas a poblaciones radicalmente diferentes tiene justificación, ya que cada una, por separado, no tendría acceso a la fuente de publicación de la otra. Como ejemplo, un estudio sobre reemplazo hormonal y enfermedad cardiovascular debería publicarse en una revista de amplia difusión, por ser un tema que interesa a muchos. Sin embargo, si se publican los resultados en una revista especializada de cardiología, los endocrinólogos y los ginecólogos, probablemente, tendrán menos oportunidad para enterarse, por lo cual se justificaría una segunda publicación en otra revista. El ideal está en que se busque la revista de mejor llegada al grupo del cual se considera que se beneficiará con la información o que más la necesita.
- Los resultados de una investigación que tenga claro interés local pueden ser publicados en una revista local para llegar a los lectores del área, y, generalmente, en su idioma; sin embargo, los resultados pueden tener también interés universal, y puede ser lícito buscar una publicación internacional (habitualmente, en inglés), y luego se obtienen los permisos para publicación adicional en la revista local.

Debe tenerse en cuenta que en todos los casos las publicaciones sobre el mismo estudio deben tratarse con criterios uniformes de autoría. No es adecuado, por ejemplo, que en un estudio sobre factores psicológicos y educativos en el manejo del asma bronquial sea el neumólogo el primer autor para la publicación en la revista de neumología, ni que el psiquiatra lo sea para la publicación en la de psiquiatría ni que la terapista respiratoria aparezca en primer lugar para el congreso de terapistas respiratorias.

En una revisión reciente, Huston[11] encontró, en búsqueda de la literatura, 20 artículos publicados y varias presentaciones no publicadas sobre experimentos clínicos aleatorizados (ECA) con risperidona, un importante antipsicótico nuevo para el tratamiento de la esquizofrenia. Sin embargo, solo se habían hecho dos experimentos clínicos grandes y algunos pequeños. Se trata de una desagregación de los mismos datos, cambios de autoría, citas múltiples y publicaciones simultáneas en diferentes revistas.

La autoría injustificada debe verse como la apropiación indebida del trabajo de otros, o la usurpación de derechos ajenos o el uso de recursos falsos para obtener beneficios. McCutchen[12] propone que se consideren siempre actos no éticos (hurto) y que se traten como tales.

Autocitación o citas injustificadas o inapropiadas

Una parte importante de una publicación son las referencias cruzadas a otras publicaciones, que sirven para dar apoyo al texto, para reconocer el aporte de otros autores o para referir al lector a otras fuentes. Cuántas veces se ha citado un estudio es una parte importante de la bibliometría, o evaluación del impacto que ha causado un artículo. Y cuantas más veces sea citado un estudio, más alta será la calificación que se hace de él en diversos índices, y más alta la retribución académica y, en ocasiones, económica que pueda derivarse del artículo.

Este beneficio de ser citado puede llevar, sin embargo, a una costumbre inapropiada, la cual va haciendo carrera, de autocitarse sin justificación o sin fundamento. Si un autor hizo una revisión de tema, y cita al generador original de un concepto, es completamente inapropiado que en un artículo posterior se cite a sí mismo por haber mencionado el concepto, pues originalmente la autoría intelectual del concepto no es suya. Se falta a la verdad al citar una fuente no original, y, posiblemente, haya una falta ética al aumentar artificialmente las citas del propio artículo[13].

Puede ser perfectamente lógico, justificado y ético si se es el generador original de la evi-

dencia y si es necesario citarse. En ocasiones estará justificado, como lo mencionan Costas *et al*.: "Dada la naturaleza acumulativa de la producción de nuevo conocimiento, las autocitas constituyen una parte natural del proceso de comunicación"[14].

Una forma de autocitación, que puede en ocasiones ser inapropiada, es citar textos propios sin identificación de su origen, o citarse innecesariamente. Aunque se sea el autor de un texto, para citar sin identificación de origen deben tenerse en cuenta aspectos como la justificación para hacerlo y los permisos que pueda ser necesario tener por parte del editor de la revista o del libro en los que esté la publicación original. Claramente, podría configurarse plagio en cualquiera de las dos situaciones: citas sin los permisos adecuados o autocitas innecesarias, referenciadas, y que pueden tener el propósito inadecuado de aumentar las citas del texto original.

CAUSAS DE LOS PROBLEMAS CON LA AUTORÍA

La autoría inmerecida concede o quita beneficios, según sea el caso. Se facilita con ella un aumento exagerado en los "méritos" de quienes son autores inmerecidos, lo cual les permite tener acceso, sin justificación, a más beneficios; esto les permite quitar oportunidades o competir deslealmente con quienes son honestos; ante la necesidad de elección entre dos candidatos para una posición para una distinción o para un ascenso, el usurpador de autoría puede ser el beneficiado.

Se han identificado varias razones para los problemas de autoría; la mayor parte de ellos son prevenibles, y *todos* son corregibles.

Falta de definición del concepto de autoría

Durante muchos años, se hizo investigación y se publicó sin que hubiese una definición generalizada de lo que significaba autoría. Con frecuencia se interpretaba de manera diversa el concepto, y se pensaba que podía ser autor quien hubiera proporcionado a sujetos para el estudio, quien financiaba parcial o totalmente el proyecto, quien hacía las pruebas diagnósticas, el jefe del servicio, quien originalmente había tenido la idea del proyecto, o, incluso, quien era considerado un experto en el tema.

Desde hace varios años se han hecho esfuerzos para lograr criterios estandarizados en la definición de autoría, se ha buscado su difusión y se han implementado medidas para garantizar, hasta cierto punto, su aplicación.

Falta de acuerdo previo entre investigadores

Aun con criterios estrictos, e, incluso, con la amplia difusión que estos han tenido, sigue habiendo dificultades en el orden de aparición y en la decisión de quiénes pueden y deben ser autores. Una razón importante para ello es la falta de acuerdo entre los autores por la omisión de un importante paso inicial en todo proyecto de investigación: no definir de manera explícita quiénes serán autores, cuáles serán las contribuciones de cada uno, cuál será el orden en el que figurarán y cuáles serán las responsabilidades de cada uno en relación con el proyecto.

AGRADECIMIENTOS Y RECONOCIMIENTOS

En la decisión de autoría, no pocas veces el grupo se encuentra con la necesidad de reconocer a personas que colaboraron de manera cercana, pero que no cumplen los criterios de autoría. Así como se debe ser justo en la decisión sobre autoría y sobre el orden de los autores, se debe buscar la justa retribución a quienes apoyaron el estudio. Los agradecimientos deben ser explícitos, en forma de nota de pie de página o al final de la publicación para quienes contribuyeron de alguna manera que no justifica autoría. O los reconocimientos a quienes en el cumplimiento de sus labores hicieron posible alguna fase del estudio.

Como ejemplos para reconocimientos, cabe mencionar a los radiólogos de un estudio de características operativas de la ecografía en el diagnóstico de apendicitis, que leyeron cada estudio dos veces, en tiempos diferentes, de manera ciega y llenando un detallado cuestionario; también, el equipo de diálisis a cargo de los pacientes en un estudio de enfermedad coronaria en insuficientes renales en hemodiálisis; los

coordinadores de investigación y los monitores, cuya misión administrativa hizo posible el estudio, pero que no tuvieron aportes científicos aceptados como autoría.

En un estudio multicéntrico, en el que se adopta un nombre colectivo, es apropiado agregar una lista de los colaboradores en todos los centros, y agradecer su esfuerzo. En estudios en los que el profesional necesita salirse de su rutina y dedicar esfuerzos especiales a la evaluación de los sujetos del estudio, o a su cuidado, aun cuando se está cumpliendo con el trabajo habitual, el esfuerzo extraordinario merece agradecimiento. Tal es el caso de un estudio de valores normales de espirometría, en el que las técnicas deben trabajar horas extras o en horarios extendidos o diferentes de los habituales para dar cabida a los sujetos del estudio. En fin, hay muchas posibles formas de colaboración por las cuales se debe agradecer. El mecanismo para hacer esto, y el cual ya se mencionó, no debe ser reemplazado con el fin de ofrecer, por gratitud, una autoría inmerecida.

ESTUDIOS SOBRE AUTORÍA

Las dificultades para definir la autoría y las diversas posibilidades de inadecuada asignación de autoría han preocupado a la comunidad científica (en particular, a lo largo de la última década), y se pueden encontrar varios estudios que se han publicado sobre el tema. Algunos, incluso, han hallado relación entre la necesidad de retractar artículos y errores en la definición de autoría[15-17].

Se ha encontrado, en general, una alta prevalencia de la autoría honoraria o de la autoría fantasma, con diferencias según el tipo de estudio. Se llevó a cabo en 2008[18] un estudio transversal en 6 revistas médicas de alto impacto, entre los autores de 896 artículos de investigación, de revisión y editoriales. En los datos del 70,3 % de los autores que respondieron la encuesta, se halló que la prevalencia de la autoría honoraria o fantasma fue del 21,0 % (IC95 % 18,0 %-24,3 %). De los artículos evaluados, el 17,6 % tuvieron autoría honoraria (IC95 % 14,6 %-21,0 %), y el 7,9 % tuvieron autores fantasma (IC95 % 6,0 %-10,3 %). Según el tipo de publicación, el 25,0 % de los reportes de investigación original tuvieron autoría honoraria, al igual que el 15,0 % de las revisiones y el 12,0 % de los editoriales, y la autoría fantasma fue del 11,9 %, el 6,0 % y el 5,3 %, respectivamente. Tales datos contrastan con un estudio similar, en las mismas revistas: en el estudio actual, se encontró una reducción significativa del 29,2 % ($p = 0{,}004$) en la prevalencia de autoría honoraria, fantasma o ambas, así como una reducción no significativa del 19,3 % ($p = 0{,}439$) en la autoría honoraria y una disminución significativa del 11,5 % ($p = 0{,}023$) en la autoría fantasma.

Otro estudio, publicado en JAMA en 1998[19], evaluó la presencia de autores honorarios y fantasma en 6 importantes revistas revisadas por pares, con una tasa de respuesta del 69,0 % (809 autores). En total, el 19,0 % de los estudios tenía autoría honoraria (rango 11 %-25 %); el 11,0 %, autoría fantasma (rango 7 %-16 %), y el 2,0 %, las dos. La prevalencia de la autoría honoraria fue mayor en los artículos de investigación (RRI 1,8; IC95 % 1,2-2,6) y la prevalencia de la autoría fantasma tuvo mayor probabilidad en los artículos de revisión (RRI 4,2; IC95 % 1,5-13,5).

Algunos estudios muestran que la probabilidad de autoría inmerecida es mayor cuanto mayor sea el número de autores del artículo. En una encuesta[20] entre autores de artículos con dos o más autores, publicados en tres revistas de farmacia y farmacoterapia en 2009, se obtuvo respuesta del 24,9 % de los encuestados. La prevalencia de la autoría honoraria fue más común en artículos de investigación original que en artículos de revisión, y hubo más autores honorarios entre los del mayor número de autores (promedio 4,9 *vs*. 3,7; $p = 0{,}002$). La proporción de artículos con autoría honoraria fue del 1,9 % para artículos con menos de 4 autores; del 25 %, para los de 4-5 autores, y del 29,4 %, para los que tenían más de 5 autores ($p = 0{,}001$).

En busca de razones para la autoría inmerecida, se han hecho varios estudios. Uno de ellos, hecho en autores de 275 artículos cuyo autor principal tuviera su residencia en Estados Unidos[21], evaluó el cumplimiento de criterios de autoría ya establecidos (porcentaje de respuesta del 72 %). Se encontró que la contribución, al menos en tres de los cuatro criterios más ampliamente aceptados (diseño de la investigación, recolección de los datos, análisis de los datos o preparación del manuscrito) había sido variable: del 99 % en el caso del primer autor, del 75 % en el segundo autor, del 48 % en el tercero y del 25 % en el cuarto. La correlación entre la

posición como autor y la contribución fue aceptable ($r = 0{,}69$; $p = 0{,}001$). Los coautores fueron citados en orden decreciente de contribución en el 70 % de los artículos; el último autor fue el segundo mayor contribuyente en el 10 % de los artículos con 3 o más autores.

La incidencia de coautores no deseados por sus compañeros pasó de ser del 9,0 % en los artículos con tres autores a ser del 30 % en los de más de seis autores (promedio del 17 %; $r = 0{,}97$; $p = 0{,}001$). En los casos en los cuales había autores no deseados, se mencionaron como principales causas para su inclusión: sentido de obligación hacia el invitado o temor por las posibles consecuencias de no incluirlo, en el 40 % de los artículos; en el 29,0 % de los estudios se citó como única contribución el reclutamiento de sujetos para el estudio.

Como un refuerzo al concepto de la necesidad de establecer los criterios de autoría, el mismo estudio mostró que el número de coautores fue menor cuando la decisión sobre la autoría se tomó al inicio del proyecto (promedio 3,9) que cuando se hizo al final de este (5,4; $p = 0{,}02$). Finalmente, la frecuencia de casos que aseguraron sentir que tenían coautores indeseados fue también menor al comparar los estudios donde la decisión se tomó al comienzo o al final (23 % *vs*. 47 %; $p = 0{,}01$).

Si bien han sido propuestos varios grupos de criterios sobre autoría, hay poca difusión y poco conocimiento al respecto, y se aplican solo en un número bajo de casos. En un estudio citado por Horton[22], se enviaron cuestionarios a 1005 estudiantes de programas de posgrado. Entre las 324 respuestas que se obtuvieron (el bajo porcentaje de respuestas lleva a pensar dos veces sobre su representatividad y sobre la potencialidad de sesgos), en más de la mitad de los casos se aceptaba que se tenía un "derecho adquirido" para ser autor sobre cualquier estudio realizado en un laboratorio solamente por el hecho de ser el director del laboratorio. El 44 % de los encuestados estaban de acuerdo en ofrecer autoría a quienes habían obtenido la financiación. La quinta parte de ellos consideraba que alguna vez habían sido injustamente excluidos de estudios en los que sí cumplían los criterios de autoría, y en el 38 % de los casos aseguraban sentir que al menos uno de sus compañeros de autoría no merecía serlo. Finalmente, un tercio de los encuestados aceptó estar dispuesto a citar inmerecidamente a alguien como autor, si eso ayudaba a la publicación.

Slone[1,5], menciona que la respuesta más frecuente entre académicos a la pregunta de cuáles serían las razones por las que citarían inmerecidamente a alguien como coautor fue para obtener promoción académica. Dado que se reconoce como causal muy importante de autoría inmerecida la presión académica para tener publicaciones, se ha buscado promover cambios en las reglas institucionales. La NLM, en Washington, cambió su formato de citas, y pasó de mencionar solo a los primeros seis autores a presentar la lista de los primeros 24 y, en todos los casos, al último autor.

PROPUESTAS DE SOLUCIONES

En el mundo académico se habla de la expresión ***publicar o perecer*** (***publish or perish***), que proviene de la urgente necesidad de tener alto número de publicaciones, por las razones ya comentadas[23]. Cabe esperar que el número de publicaciones siga aumentando, como ha sido hasta ahora, y que el número de coautores también aumente. Por supuesto, los problemas para la definición de autoría irán creciendo de manera simultánea con las publicaciones.

COMITÉ INTERNACIONAL DE EDITORES DE REVISTAS MÉDICAS (ICMJE)

El Comité Internacional de Editores de Revistas Médicas (ICMJE, por las siglas en inglés de International Committee of Medical Journal Editors) es un grupo de editores de revistas médicas generales, encargados de escribir las *Recomendaciones para la conducción, reporte, edición y publicación de obras científicas en revistas médicas*, cuya última versión fue publicada en 2013[10]. Su actividad se inició en 1985, durante una reunión en Vancouver, Canadá, entre un grupo de editores de revistas médicas que querían evaluar el problema creciente, surgido con la responsabilidad en las publicaciones, y primaria, pero no únicamente, relacionado con la autoría.

Las recomendaciones fueron desarrolladas para promover la mejor práctica y el mejor cumplimiento de estándares éticos en las publicaciones científicas en revistas médicas,

aunque hay otros propósitos adicionales. Las recomendaciones no tienen fuerza de ley, pero el mundo académico las ha aceptado cada vez más, y aunque se han propuesto primariamente para las revistas pertenecientes al ICMJE, se sugiere su uso para todas las publicaciones.

DEFINICIÓN DE CRITERIOS DE AUTORÍA

El ICMJE refuerza la importancia de la autoría porque confiere crédito y por sus importantes implicaciones académicas, sociales y financieras. Adicionalmente, implica responsabilidad y rendición de cuentas por el trabajo publicado. El grupo busca que haya claridad en los criterios para establecer la autoría, como una forma de proteger los derechos de los autores; pero al mismo tiempo busca que los autores tengan consciencia de su papel y asuman su responsabilidad y entiendan que deben responder por lo que se publicó.

Como estrategia para promover el respeto a la autoría, el ICMJE propone, además de los criterios básicos de autoría, que las revistas pidan y publiquen información sobre las contribuciones que cada autor haya hecho en el estudio. Se propone, así mismo, que las revistas desarrollen e implementen políticas de contribuciones, las cuales permitan identificar expresamente a quienes se harán responsables por el trabajo como un todo.

RECOMENDACIONES ESPECÍFICAS SOBRE AUTORÍA

La propuesta del ICMJE tiene cuatro criterios (tabla 37.2) en los que debe basarse la decisión de autoría, y que deben cumplirse en conjunto (no solo algunos, sino todos ellos)[10]:

Los criterios, que deben ser cumplidos en su totalidad, comprenden las fases más importantes de un trabajo de investigación:

1. *Contribución sustancial a la concepción o el diseño del trabajo*; o la adquisición, el análisis o la interpretación de datos para el trabajo: Es de suma importancia que la contribución haya sido sustancial, y que el trabajo en adquisición, análisis o interpretación de los datos se acompañe de las fases siguientes. Adicionalmente:
2. *Escritura del borrador del trabajo, o revisión crítica del contenido intelectualmente importante:* Las contribuciones tanto en la escritura como en la revisión crítica (puede ser solamente una de las dos) deben ser sustanciales, y la revisiones deben haber sido de aspectos de relevancia intelectual, y no solo de revisión de estilo. Adicionalmente:
3. *Aprobación final de la versión que se publicará:* Se trata, nuevamente, de la aprobación luego de revisión crítica, y significa que se asegura, por parte de cada autor, que se han revisado en profundidad todos los aspectos del manuscrito, y que se asume la responsabilidad futura por todos ellos, al tiempo que se garantiza su veracidad. Adicionalmente:
4. Aceptación de que se rendirán cuentas y de que se es responsable de todos los aspectos del trabajo, y así se asegura que cada aspecto relacionado con la precisión o la integridad de cualquier parte del trabajo ha sido investigado y resuelto de manera apropiada.

Además de ser responsable por las partes del trabajo que cada autor haya realizado, este debe poder identificar a los coautores responsables de otras partes específicas del trabajo. Adicionalmente, se debe tener confianza sobre la integridad de las contribuciones de sus coautores.

Todos aquellos designados como autores deben cumplir los cuatro criterios de autoría, y quienes los cumplan deben ser identificados como autores. Quienes no cumplan los cuatro criterios deben recibir reconocimiento, pero en las secciones de agradecimiento o de reconocimientos.

Tabla 37.2 Criterios de autoría (ICMJE)

1. Contribución sustancial en la concepción, el diseño, la recolección de datos, el análisis o la interpretación, **Y**
2. Escritura o revisión crítica del contenido intelectualmente importante, **Y**
3. Aprobación final de la versión que se publicará, **Y**
4. Aceptación de responsabilidad y de que se rendirán cuentas.

COMPROMISO CON QUIENES CONTRIBUYEN

El ICMJE va aún más allá de los criterios anteriormente expuestos, y expresa que si una

persona cumple el primer criterio, debe recibir la oportunidad de participar en la escritura, la revisión y la aprobación final del manuscrito. No se trata solo de quién cumplió: se trata de que se respete el derecho que tiene quien participa en la concepción, la conducción y el análisis a participar en las fases finales.

Idealmente, en la propuesta del ICMJE quienes conducen el estudio tienen la responsabilidad de identificar a quienes cumplen los criterios, y también, la de hacer esto, idealmente, en las etapas de planeación del trabajo, así como la de hacer ajustes sobre la marcha cuando sea necesario.

Debe recordarse que la responsabilidad de la decisión correcta y justa sobre quiénes figuran como autores es de ellos mismos, y no del editor de la revista a la que se envíe. Tampoco es responsabilidad de la revista determinar quiénes califican como autores, ni arbitrar conflictos de autoría. En caso de desacuerdos, corresponde a la institución o las instituciones donde se hizo el estudio resolverlos, no a los editores de la revista[10].

En algunos casos, en estudios multicéntricos con un gran número de autores, puede recurrirse a un nombre del grupo, que aparecerá entre los autores, y en un apéndice debe ofrecerse el listado completo de quienes cumplan los criterios.

Si un autor no cumple los cuatro requisitos, debe recibir un reconocimiento, llamado en algunos casos ***contribución sin autoría***. Se trata, por ejemplo, de quienes han conseguido la financiación, o han llevado a cabo la supervisión general del grupo de investigación, o han brindado apoyo administrativo general, o asistencia en la escritura, en la edición técnica, en la revisión de estilo y en la corrección de pruebas[10]. A estas personas se les podrá agradecer individualmente o como grupo, y sus contribuciones deben especificarse, pero no deben ser citadas como autores.

Vale la pena evaluar la contribución individual de cada uno de los autores; particularmente, en aspectos como los siguientes, que pueden usarse como lista de verificación: supervisión de la labor de los coautores, reclutamiento de coautores, reclutamiento de sujetos de estudio, análisis y explicación de datos, búsqueda de la literatura, análisis y explicación de esta, revisión del manuscrito, comunicación con el editor de la revista, firma de la declaración de transferencia de derechos de autor, concepción y diseño del trabajo, recolección de datos, obtención de financiación o de material de soporte, realización del análisis estadístico, escritura del manuscrito o de parte de él, aprobación del manuscrito antes de su envío a la revista, revisión del manuscrito con cambios importantes en su contenido, corrección de pruebas (galeradas), etc.

OTRAS SOLUCIONES PROPUESTAS

La propuesta inicial del ICMJE es la base para solucionar el problema y prevenir su empeoramiento, pero debe acompañarse de medidas adicionales: difusión y exigencias universales son las principales.

Hay un riesgo enorme en la adopción de medidas locales, ya que pueden ser muy diferentes entre sitios, y aun cuando puede facilitarse su aceptación, van a representar un desequilibrio entre diferentes sitios, lo cual generará injusticias. Por ejemplo, si un departamento decide aceptar que todos sus miembros pueden ser coautores de todos los trabajos, aun sin el cumplimiento estricto de los criterios de autoría, generará una situación de desigualdad con los miembros de otros departamentos del mismo hospital o universidad y, por supuesto, con las demás instituciones de la región y del país. Ante la necesidad de selección por méritos, entre los que se considere el número de publicaciones, los generosos "donadores" de autoría aparecerán inmerecidamente por delante de quienes solo presentan las publicaciones en las que sí merecen autoría.

Algunas universidades han puesto límite al número de artículos que tendrán en cuenta para promociones académicas (diferente según el nivel en el escalafón de los profesores asistentes, los asociados o los titulares). También se ha llegado, en algunos casos, a fijar de antemano políticas sobre los criterios para establecer el orden adecuado de autoría.

Como los editores no deben servir como policías ni están dispuestos a hacerlo —según anotan Kassirer y Angell (editores, a su vez)[8]—, todavía debe seguirse confiando en la honestidad y en la rectitud de los científicos que buscan publicar un trabajo.

Deben ofrecerse los mecanismos para que su honestidad y su rectitud se vean ayudadas por un grupo de criterios lógicos y aplicables, ampliamente conocidos y aceptados. Se necesita

establecer reglas claras de autoría, su difusión amplia y, principalmente, la educación al público general, a los médicos, a los administradores académicos y a los investigadores, respecto a la necesidad de criterios de autoría, sobre los criterios mismos, sobre su aplicación y sobre los riesgos y los daños derivados de su incumplimiento.

Algunas revistas incluyen en sus requisitos la necesidad de una declaración firmada por cada uno de los autores, en la que establezcan que cumplen los cuatro criterios básicos. Igualmente, se propone que cada autor especifique si forma parte de otro estudio similar o del mismo con diferente presentación o con diferentes autores[24].

MECANISMOS PARA FACILITAR LA AUTORREGULACIÓN Y EL CONTROL

Como parte de las propuestas presentadas por el ICMJE, se han incluido ayudas que les permiten a los autores tener certeza de que se cumplen los requisitos, a la vez que permiten mayor transparencia en el proceso de evaluación.

Entre las propuestas está[10] el envío de una carta de presentación del artículo, que especifique claramente si se han hecho o no publicaciones previas o si el trabajo ha sido enviado a otra revista. En caso de que haya una publicación previa, así debe especificarse y enviarse copia del material.

También debe haber una clara mención de las relaciones financieras, laborales o de otra índole que puedan significar conflictos de interés.

Algunas revistas exigen una declaración formal explícita de la contribución específica de cada uno de los autores. Otras revistas piden que haya una carta de sometimiento con una declaración de que el manuscrito ha sido leído, revisado y aprobado por todos los autores, de que se han cumplido los requisitos de autoría y de que cada autor cree que el artículo representa un trabajo honesto.

NORMAS ESPECÍFICAS ACERCA DE PUBLICACIONES EN INTERNET

La acogida cada vez mayor que tiene la publicación en Internet, así como la facilidad para hacerlo, ha traído problemas nuevos, los cuales evolucionan a medida que se desarrollan las redes[25].

Uno de los riesgos principales es la publicación fácil, sin criterios de calidad, con amplio número de lectores, gracias a los espacios disponibles. La falta de necesidad de revisión por pares, la ausencia de una competencia real y la libertad casi absoluta llevan a que haya publicaciones más fáciles y de menor calidad en la red que en el papel.

Los derechos de los autores son más fáciles de violar, bien por la facilidad para copiar y alterar los textos, bien por la libertad o la facilidad para la copia y la edición de los textos. Las precauciones deben ser infinitas, y los problemas tan solo han empezado a aparecer[25,26].

La rápida y enorme disponibilidad de información en Internet (y, particularmente, la opción de acceso a la literatura médica) ha producido nuevos problemas a médicos, académicos y estudiantes: control de calidad, dificultad para establecer los componentes básicos, falta de guías estándar para citas, corta vida de las direcciones de internet y una gran amenaza al concepto de propiedad intelectual, por la disponibilidad de información completa (tablas, cifras, gráficos, datos completos).

La publicación electrónica ha empezado a reemplazar la publicación en papel, y debe, por lo tanto, cumplir con los mismos requisitos, si bien hay que establecer algunas normas especiales. El ICMJE[10] propone que se sigan las mismas políticas establecidas para las publicaciones en papel.

CONCLUSIONES

Es necesario establecer con claridad los requisitos que deben cumplirse para poder reclamar la autoría de un trabajo de investigación, no solamente para evitar inclusiones inadecuadas e injustificadas, sino para asegurar que no se cometan omisiones. De la misma forma, es importante que haya una definición precisa del orden adecuado de los autores, que indique fielmente cuánto contribuyó cada uno al trabajo intelectual y cuánto aportó cada uno al resultado final.

Deben definirse de manera estricta el número y el orden correcto de autores, pues la autoría significa tener la responsabilidad sobre el contenido de la investigación y sobre su difusión;

adicionalmente, significa un compromiso con la comunidad científica internacional en cuanto al aseguramiento de un manejo impecable de los datos, no solo por la veracidad y por el rigor científico y metodológico, sino por el respeto a los principios éticos y legales.

Deben tenerse en cuenta también los agradecimientos y los reconocimientos a quienes los merezcan, en la proporción justa que corresponda a su participación.

Finalmente, debe garantizarse que se aclaran las condiciones en las que puede hacerse una publicación repetida o varias publicaciones sobre el mismo proyecto, sin que ello signifique que se exagera la actividad con múltiples presentaciones injustificadas de un mismo tema.

Aunque es fundamental promover la promulgación de normas para vigilar la asignación de la autoría y tratar de que haya definición (con participación) de reglas de juego y difusión amplia, debe aceptarse como un principio ideal —de mayor importancia— que el mundo científico debe seguir teniendo como base la honestidad y la rectitud de los investigadores y los científicos.

Cabe una responsabilidad muy especial a las universidades, que, además, deben formar a sus estudiantes con los conocimientos, las destrezas y las actitudes necesarios para la práctica adecuada de buena ciencia, para el ejercicio honesto de su profesión y para la producción en investigación impecable desde cualquier punto de vista.

Referencias

1. Fletcher SW, Fletcher RH. Responsibilities of medical journals to readers. J Intern Med. 1992;232:223-8.
2. Hamblet JL. Ethical issues in publication. Semin Perioper Nurs. 1996;5:102-7.
3. Kapoor VK. Polyauthoritis giftosa. Lancet. 1995;346:1039.
4. Swiss Academies of Arts and Sciences (SAMW). Authorship in scientific publications. Analysis and recommendations Scientific Integrity Committee of the Swiss Academies of Arts and Sciences. Berna, Suiza: SAMW; 2013.
5. Abbenhaim L, Moride Y, Brenot L. Appetite-suppressant drugs and the risk of primary pulmonary hypertension. N Engl J Med. 1996;335:609-16.
6. Kassirer JP, Angell M. Financial conflicts of interest in biomedical research. N Engl J Med. 1993;329:570-1.
7. Wadman M. Drug company links test journal's policy. Nature. 1996;383:6.
8. Kassirer JP, Angell M. The journal's policy on cost-effectiveness analyses. N Engl J Med. 1994;331:669-70.
9. Shapiro DW, Wenger NS, Shapiro MF. The contributions of authors to multiauthored biomedical research papers. JAMA. 1994;271:438-42.
10. International Committee of Medical Journal Editors (ICMJE). Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals [internet]. 2014 [citado 2015 mar. 13]. Disponible en: http://www.icmje.org/icmje-recommendations.pdf
11. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. Lancet. 1996;347:1024-6.
12. Mccutchen CW. Authorship disputes. Science. 1995;270:1553--7.
13. Fowler JH, Aksnes DW. Does self-citation pay? Scientometrics. 2007;72:427-37.
14. Costas R, van Leeuwen TN, Bordons M. Self-citations at the meso and individual levels: effects of different calculation methods. Scientometrics. 2010;82:517-37.
15. Eastwood S, Derish P, Leash E, et al. Ethical issues in biomedical research: perceptions and practices of postdoctoral research fellows responding to a survey. SciEng Ethics. 1996;2:89-114.
16. Hoen WP, Walvoort HC, Overbeke AJ. What are the factors determining authorship and the order of the authors' names? A study among authors of the Nederlands Tijdschriftvoor Geneeskunde (Dutch Journal of Medicine). JAMA. 1998;280:217-8.
17. Wager E, Williams P. Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. J Med Ethics. 2011;37:567-70.
18. Wislar JS, Flanagin A, Fontanarosa PB, et al. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ. 2011;343:d6128.
19. Flanagin A, Carey LA, Fontanarosa PB, et al.Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA. 1998;280:222-4.
20. Dotson B, Slaughter RL. Prevalence of articles with

honorary and ghost authors in three pharmacy journals. Am J Health Syst Pharm. 2011;68:1730-4.
21. Slone RM. Coauthors contributions to major papers published in the AJR: Frequency of undeserved coauthorship. AJR Am J Roentgenol. 1996;167:571-9.
22. Horton R. The unmasked carnival of science. Lancet. 1998;351:688-9.
23. Rennie D, Yank V, Emanual L. When authorship fails: a proposal to make contributors accountable. JAMA. 1997;278:579-85.
24. Guarding the guardians: research on editorial peer review. Selected proceedings from the First International Congress on Peer Review in Biomedical Publication. May 10-12, 1989, Chicago, Ill. JAMA. 1990;263:1317-441.
25. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector etviewor--Let the reader and viewer beware. JAMA. 1997;277:1244-5.
26. Litynski GS. Click c@refully before you quote: citing internet-based sources. JSLS. 1999;3:235-9.

CAPÍTULO

38

Informática biomédica y epidemiología clínica

FERNANDO SUÁREZ OBANDO

INTRODUCCIÓN

La informática biomédica comprende los aspectos teóricos y prácticos del manejo de la información y de la comunicación relacionados con el desarrollo y la aplicación de los procesos de atención sanitaria[1].

Se define como la ciencia que utiliza herramientas informáticas y analíticas para facilitar la administración y el uso de información que apoyen los procesos de investigación clínica y la práctica del profesional en salud[2].

El vertiginoso desarrollo de la informática biomédica ha convertido esta disciplina científica en un elemento indispensable de la atención y de la investigación clínicas[3], y sus áreas de influencia comprenden desde el desarrollo de la historia clínica electrónica (HCE) hasta la atención de pacientes a través de redes de telemedicina, incluyendo diversos aspectos de importancia médica, como el procesamiento digital de imágenes diagnósticas, los sistemas inteligentes de apoyo diagnóstico, el perfeccionamiento de la terminología clínica, el mejoramiento en la clasificación de las enfermedades y el desarrollo de nuevas tecnologías que enlazan la bioinformática con la atención clínica[4-7].

La informática biomédica permite potenciar el impacto de la medicina traslacional y ampliar el horizonte de la epidemiología clínica, al ofrecer las herramientas que integran efectivamente datos de salud pública, datos históricos, guías de manejo y literatura científica a la información clínica del paciente, con el fin de optimizar la gestión del conocimiento y el análisis de datos sanitarios[8,9].

Desde sus inicios, la informática biomédica, al desarrollar las herramientas necesarias para gestionar grandes volúmenes de información, ha revolucionado el modo de ejercer la medicina integrando dentro de la práctica clínica las tecnologías de la comunicación y de la información (TIC)[10].

FUNDAMENTOS DE LA INFORMÁTICA BIOMÉDICA

La informática biomédica es un campo interdisciplinario y científico que estudia y analiza diversos usos de los datos biomédicos en el campo de la investigación médico-científica, en la generación de soluciones a problemas de la práctica asistencial y en el desarrollo de herramientas informáticas y analíticas para la toma de decisiones clínicas.

El surgimiento de esta disciplina se correlaciona directamente con la invención y la diseminación de la tecnología computacional y con el desarrollo de tecnologías de la comunicación basadas en sistemas digitales[11].

La aplicación de las tecnologías digitales al ámbito de la salud se ha encaminado a través de la informática biomédica, que se cimienta en tres pilares fundamentales para su desarrollo: 1) El uso de computadores debe ser en beneficio de las personas, y no en reemplazo de ellas. 2) El recurso tecnológico debe ofrecer a las personas una novedad, una mejora o una trasformación en sus actividades que optimice sus acciones y que no sea la repetición de un recurso ya conocido por el usuario. 3) Debe existir una interacción entre el recurso tecnológico y el usuario, una interacción sobre la cual los resultados no sean predecibles; es decir, que el resultado final del uso de un recurso tecnológico solo se podrá observar en la medida en la que dicho recurso se utilice.

Estos tres principios implican que la informática biomédica se trata de un desarrollo tecnológico dirigido a las personas, que mejora sus capacidades y que define el progreso de la disciplina al implementar procesos de investigación, tanto en el desarrollo del recurso, como en su implementación, en las pruebas de concepto y en el uso cotidiano del recurso; para detectar errores o situaciones problemáticas cuya solución permita el progreso de las tecnologías, un aspecto que solo es posible al evaluar la interacción entre usuarios y tecnología[12].

Los fundamentos de la informática biomédica, a su vez, encuentran su asidero en la interacción entre el cuidado del paciente, la prestación de servicios y el sistema de salud en el contexto de la gestión de la información y su comunicación; es decir, la captura, la entrega, la transmisión y el uso de datos útiles para la toma de decisiones clínicas.

La informática biomédica se vale, principalmente, de la gestión de la información en salud como elemento esencial para la prestación eficaz y eficiente de la asistencia sanitaria[13,14].

INTERACCIÓN DE LA INFORMÁTICA BIOMÉDICA EN LA EPIDEMIOLOGÍA CLÍNICA

Más allá del uso de computadores en un hospital, la informática biomédica es fundamental para el desarrollo de la medicina, pues la interacción entre los usuarios de la información sanitaria y las herramientas tecnológicas disponibles, genera escenarios que contribuyen al desarrollo de la atención médica en nuevos entornos (por ejemplo, Internet, telemedicina, virtualización) y revela la necesidad de decantar la información en acciones clínicas concretas administrando diversas fuentes de datos y analizando el impacto que el acceso a la información sanitaria tiene sobre el sistema de salud y sobre los desenlaces clínicos.

En ese sentido la informática biomédica sirve a la epidemiología clínica, tanto como complemento de su actividad investigativa como de herramienta para la recolección y análisis de datos. En la **tabla 38.1** se describen algunos de los dominios de desarrollo y actividades investigativas de la informática biomédica, que se correlacionan directamente con el proceder de la epidemiología clínica, interacción que da como fruto la optimización de procesos de investigación, así como la aplicación de metodología investigativa en el desarrollo de soluciones tecnológicas.

Esta interacción entre la informática biomédica y la epidemiología clínica se da naturalmente, pues herramientas de informática, tales como computadores, tabletas, celulares inteligentes, Internet, telemedicina y otras tecnologías, son parte integral del entorno cotidiano de los profesionales de la salud[16-18], lo cual implica que la investigación en torno a la atención sanitaria necesita manejar la complejidad de los datos, con la participación de representantes de diversas disciplinas que tengan la capacidad de comunicarse e intercambiar datos electrónicamente[19,20] y gestionar enormes volúmenes de información con herramientas diseñadas para los entornos de atención en salud[21,22], así como generar la capacidad necesaria para desarrollar y evaluar el impacto de los sistemas de los información y comunicación como una destreza que complemente los conocimientos de la medicina basada en la evidencia[23,24].

Por otra parte, la investigación biomédica requiere un fundamento tecnológico suficientemente robusto para soportar la demanda de soluciones a problemas biomédicos cada vez

Tabla 38.1 Procesamiento de datos, información en medicina y cuidado de la salud en interacción con la epidemiología clínica. La informática biomédica se describe en términos de dominios de conocimiento, cada uno de los cuales tiene actividades específicas de desarrollo que encuentran en la epidemiologia clínica su entorno adecuado de implementación.

Informática biomédica		
		Actividades de innovación e investigación
Dominios	Identificación y uso de la información sanitaria a través de medios electrónicos	Procesamiento sistemático de datos, información y conocimiento médicos, aplicados a la asistencia sanitaria.
		Identificación de las ventajas y las limitaciones del uso de TIC en medicina y en asistencia sanitaria.
		Identificación de datos e información de alta calidad aplicables a la solución de problemas de salud del paciente y a problemas relacionados con la gestión institucional sanitaria.
		Identificación de necesidades de desarrollo de estrategias de información, enseñanza y entrenamiento del personal de salud en habilidades para el procesamiento sistemático de datos e información.
	Soporte de decisiones diagnósticas	Uso eficiente y responsable de las herramientas de procesamiento de la información, con el fin de optimizar la práctica del personal de salud y su proceso de toma de decisiones.
	Características generales de los SIS	Operación y administración de sistemas de información sanitaria.
		Conocimiento de la arquitectura de diversos tipos de SIS; especialmente, los utilizados como sistemas de información hospitalaria y de consultorio destinados al soporte de la práctica clínica o al gerenciamiento de instituciones.
		Conocimiento de la arquitectura de diversos tipos de SIS aplicados a la salud pública.
		Uso adecuado de *software* para la documentación, el análisis y la comunicación de datos derivados de SIS.
	Estructura y diseño de la historia clínica electrónica y de los sistemas de apoyo diagnóstico	Conocimiento de la estructura, el diseño y los principios de análisis de la HCE, incluyendo nociones de calidad de los datos, conjuntos mínimos de datos y aplicaciones de la HCE en ámbitos investigativos, administrativos y de vigilancia epidemiológica.
		Conocimiento de los principios y la estructura de los principales sistemas de codificación de terminología médica, y sus aplicaciones en el ámbito de la HCE.
		Comprensión de los procesos de soporte diagnóstico basados en sistemas electrónicos y computacionales aplicados a la atención en salud, el desarrollo y la ejecución de guías de manejo y la identificación y la aplicación de rutas críticas.
		Construcción de herramientas y de sistemas de apoyo diagnóstico aplicados al manejo de pacientes, así como de otras herramientas orientadas a la administración de recursos en salud y adquisición, representación e ingeniería del conocimiento médico.
	Aplicaciones de la informática médica para la solución de problemas de salud	Conocimiento y aplicación de los desarrollos y los fundamentos de la telemática en salud.
		Conocimiento y aplicación de los fundamentos del comercio electrónico en salud.
		Conocimiento y aplicación de los desarrollos y los fundamentos del procesamiento de imágenes.
		Conocimiento y aplicación de los principios generales de Ingeniería Biomédica
		Conocimiento y aplicación de los principios generales de la robótica médica.
		Conocimiento y aplicación de los desarrollos y los fundamentos de la virtualización de estructuras anatómicas y procesos fisiológicos.
		Conocimiento y aplicación de las herramientas de la informática biomédica empleadas en el análisis bioestadístico.

Continúa

Continuación

	Interacción de la informática médica con otras disciplinas y saberes	Conocimiento y aplicación de las herramientas de la informática biomédica desarrolladas para el mejoramiento de los sistemas de salud en carreras afines a la medicina.
		Interacción y desarrollo de la informática aplicada a la enfermería.
		Conocimiento de las implicaciones éticas del desarrollo de la informática médica identificando los problemas de privacidad, confidencialidad y seguridad de los datos personales y sensibles en entornos electrónicos.
		Desarrollo de procesos adecuados para salvaguardar la información, así como identificar y plantear soluciones a dilemas éticos relacionados con la práctica de la informática médica.
	Investigación, innovación y desarrollo	Aplicación de los conocimientos de informática biomédica en el desarrollo de investigación biomédica.
		Participación en equipos interdisciplinarios de investigación que estudien el impacto de la informática biomédica en la atención en salud.
		Desarrollo de investigación conjunta con otras disciplinas como la epidemiología clínica y la salud pública.
		Participación en equipos de desarrollo de herramientas informáticas aplicadas a la salud, que tengan proyección hacia recursos tecnológicos patentables y comercializables.

Nota: SIS: sistemas de información sanitaria; HCE: historia clínica electrónica; TICs: tecnologías de la información y la comunicación.

Fuente: Adaptado de Mantas[15].

más complejos[25,26], lo que incluye sistemas de vigilancia epidemiológica y de salud pública que se alimentan de grandes volúmenes de datos, administrados a través de complejos sistemas informáticos. Esta interacción permite generar mejores modelos de predicción que, a su vez, propongan intervenciones clínicas más tempranas y eficientes[27,28].

La presencia de la informática biomédica como disciplina que interactúa con otras áreas del conocimiento médico-científico permite generar nuevos campos de investigación que abren nuevas posibilidades a la epidemiologia clínica.

Permite contar con nuevas herramientas de recolección y análisis de los datos, como, por ejemplo, la medición precisa de la costo-efectividad de programas de seguridad del paciente y la disminución de errores médicos en el contexto de la HCE, suponiendo que los errores médicos pueden controlarse con el desarrollo y la implementación de herramientas informáticas que asistan los procesos clínicos[29,30], y que la seguridad del paciente se incrementa con la introducción de diversos sistemas e innovaciones tecnológicas, utilizados para prevenir accidentes en el entorno clínico[31,32].

El desarrollo y la implementación de tecnologías digitales en el ámbito clínico implica costos e inversiones que deben ser calificados y evaluados utilizando la convergencia disciplinar entre la evaluación económica, la evaluación epidemiológica, el impacto clínico y la evaluación tecnológica.

Por otra parte, las organización de salud y los centros académicos y de investigación biotecnológica están volcando su quehacer hacia un desarrollo sostenible, lo que implica la evaluación continua de los sistemas de salud; especialmente, en el caso de la implementación de nuevas tecnologías, para lo cual se requieren herramientas que integren y administren eficientemente diversas fuentes de información[33,34].

La informática biomédica es fundamental en el contexto de la medicina traslacional, que tiene como objetivo principal la aplicación clínica de las investigaciones básicas, a través de la implementación práctica de herramientas informáticas[35,36], lo que implica, a su vez, en un futuro cercano, la integración de información genómica, eómica y metabolómica a la HCE, un paso fundamental en el desarrollo de una atención médica que incluirá plenamente el papel de la biología molecular como herramienta predictiva, terapéutica y de pronóstico dentro de la atención médica rutinaria[37,38].

La epidemiología clínica ha desarrollado las herramientas necesarias para interactuar en el proceso de optimización del uso de recursos para la salud, lo que en el mundo digital depende, en gran medida, de la adecuada implementación de recursos electrónicos[39,40]. Es decir, la gestión de la información aplicada a los sistemas de salud, mediada por la epidemiología clínica, encaminará a la medicina a solucionar problemas de cobertura y de calidad en la atención en salud siguiendo patrones de innovación tecnológica que incluyen la informática y las telecomunicaciones[41,42]. Como consecuencia, la atención clínica incluirá nuevos actores en el proceso de cuidado del paciente interactuando en un amplio espectro de desarrollo que incluirá, entre otros, los sistemas inteligentes de soporte a las decisiones clínicas y la robótica clínica[43,44].

INTERACCIÓN ENTRE EPIDEMIOLOGÍA, INFORMÁTICA BIOMÉDICA Y DOCENCIA

La preparación que el profesional de la salud debe tener para desempeñarse en el ámbito asistencial o para desarrollar investigación epidemiológica obliga, dentro del contexto de la informática y las comunicaciones, a replantear los fundamentos de su conocimiento, pues se hace indispensable que el profesional conozca las bases sobre las cuales se desarrollan las herramientas informáticas. Esto, con el fin de aplicar adecuadamente un recurso, según los contextos clínicos o investigativos.

La complejidad actual del conocimiento médico requiere el uso de estándares informáticos que permitan un diálogo preciso entre investigadores, personal de salud, pacientes y entes gubernamentales, estándares que se basan en el desarrollo de terminología médica controlada a través de modelos planteados por la informática biomédica[45,46]. Es así como los profesionales de la salud deben tener las habilidades para gestionar sistemáticamente datos, información y conocimiento, y, además, comprender la forma como el manejo de la información influye tanto en el proceso como en la calidad de la toma de decisiones[13,47]. Por tanto, comprender la naturaleza de los sistemas de información y comunicación en la prestación de la asistencia sanitaria es fundamental en la preparación del médico[15,48].

La enseñanza de la medicina y de las demás profesiones afines a la salud se verá beneficiada por la virtualización de estructuras anatómicas y de procesos fisiológicos, así como por la simulación de procesos patológicos y por procesos de enseñanza que se complementan con los modelos de aplicación clínica de la informática biomédica[49,50]. Pero se verá más beneficiada por la enseñanza de una medicina que incorpore sistemas complejos de gestión de la información digital, desarrollo de sistemas inteligentes de diagnóstico y sistemas de vigilancia epidemiológica, integrados a las historias clínicas electrónicas[51,52].

RETOS DEL USO DE LA INFORMÁTICA BIOMÉDICA

La informática biomédica utiliza el poder de la revolución tecnológica de las TIC para representar un conjunto de conocimientos con el fin de mejorar la calidad de la atención e incrementar el acceso al sistema sanitario. Sin embargo, su vertiginoso desarrollo implica también riesgos y nuevos retos[53,54], los cuales deben ser tenidos en cuenta para que el empleo de la tecnología no tenga consecuencias negativas ni para el paciente ni para el personal de salud.

Algunos de los mencionados riesgos comprenden: la despersonalización de la atención médica, por la intermediación tecnológica dominante[55,56]; la excesiva confianza en los sistemas electrónicos, que puede quitar peso al criterio médico[57,58]; el desplazamiento del personal de salud como principal fuente de información sanitaria[59,60]; la proliferación y la aceptación de supuestas soluciones a problemas clínicos basadas en sistemas electrónicos, sin que exista un soporte o una evidencia científica de su utilidad o su verdadero impacto[61,62], y el uso de herramientas informáticas en ambientes que difieren significativamente del entorno donde fueron desarrolladas; y en el cual puede haber descontextualización de las herramientas según sus objetivos, sus propósitos y sus medios de implementación[63,64].

Cualquier tecnología tiene riesgos en su implementación, y la informática biomédica no es la excepción. Uno de los mayores riesgos que corre el proceso de implementar los sistemas digitales en un sistema de salud es la excesiva confianza en los sistemas electrónicos en relación con la seguridad, la privacidad y la confidencialidad de los datos clínicos[65,66].

La excesiva confianza en los sistemas de seguridad incrementa potencialmente las probabilidades de acceso a información sensible por parte de terceros[67,68], lo que implica la búsqueda constante de un balance entre seguridad y accesibilidad lo que puede aumentar los costos de desarrollo con el fin de cumplir con parámetros de seguridad.

Por otra parte, la necesidad de mantener un contexto adecuado de usabilidad con soluciones novedosas genera el incremento en la oferta de soluciones informáticas, sin que necesariamente se cumpla en todos los casos con estándares mínimos de interoperabilidad[69,70], lo que dificulta la comunicación entre los sistemas adoptados por diversas organizaciones prestadores de salud. Esto último puede llevar a mayores riesgos de fuga y de manejo inadecuado de la información[71-74].

Además de las limitaciones propias de la tecnología, su implementación genera consecuencias negativas, que no necesariamente podían contemplarse desde las fases de diseño. Introducir un sistema informático en el ámbito clínico genera un incremento en la carga de trabajo asociada a la introducción de nuevas tecnologías[75,76], distorsiona las rutinas y las prácticas tradicionales y así genera nuevos errores en la atención clínica, mediados por la introducción de tecnologías electrónicas[77,78]; además, crea incomodidad en el personal de salud, ante la obligatoriedad de utilizar tecnologías informáticas en ambientes clínicos [79,80] y estimula cambios inesperados en las estructuras organizacionales de atención clínica tras la introducción de tecnologías informáticas[81,82].

Todos los anteriores aspectos deben tenerse en cuenta en las evaluaciones de impacto económico que se lleven a cabo sobre las tecnologías introducidas a los sistemas de salud, pues el desbalance entre la inversión económica para el desarrollo y la implementación de tecnologías en la atención clínica y los beneficios alcanzados por su implementación pueden originarse en el impacto negativo de la propia tecnología entre los usuarios[83,84].

PROYECCIÓN CONJUNTA DE LA INFORMÁTICA BIOMÉDICA Y DE LA EPIDEMIOLOGÍA CLÍNICA

La naturaleza de la informática médica le permite ser una herramienta adecuada tanto para los procesos de mejoramiento de los sistemas de salud como de la investigación aplicada[85,86].

La informática médica es lo suficientemente versátil para llevar a la práctica resultados de investigaciones y conceptos de las ciencias básicas y clínicas; o sea, es un puente entre las necesidades de la práctica médica y los planteamientos de la medicina traslacional, y que conjuga características que la hacen afín a la epidemiología clínica y a la bioestadística, disciplinas que utilizan herramientas de la informática médica en prácticamente todos sus procesos[85].

Esta interacción y esta transversalidad de saberes generan un círculo virtuoso, donde los objetivos de la epidemiología clínica se complementan con los objetivos de la informática biomédica, dado que la investigación en salud identifica los principales problemas en el área, plantea soluciones y da sustento al progreso del conocimiento con la aplicación del método científico, mientras que la informática biomédica procura la adecuada administración de la información, generar procesos de gestión del conocimiento y aplicar la ciencias de la computación en generar modelos, aplicaciones y algoritmos que optimicen los proceso de investigación, y su posterior aplicación en contextos de práctica clínica; es decir ambas áreas complementan, retroalimentan y tienden a perfeccionar los procesos inherentes a cada disciplina.

Por otra parte, tanto la epidemiología clínica como la informática biomédica son disciplinas transversales en todas las especialidades médico-quirúrgicas, e influyen en los procesos docentes, investigativos y académicos buscan la integridad de los procesos de innovación y desarrollo utilizando el lenguaje común de la ciencia

y coinciden el objetivo de mejorar las condiciones de salud desde perspectivas diversas, como la salud pública, la atención primaria en salud o el análisis de imágenes diagnósticas. Tales características les permite a ambas disciplinas interactuar con todas las áreas relacionadas con la prestación de servicios sanitarios y comunicarse fácilmente con otras áreas del conocimiento, y fortalece así la interdisciplinariedad.

Finalmente, la práctica de la medicina moderna exige con mayor énfasis que los resultados de las investigaciones generen beneficios para los pacientes[87], aspecto crítico en relación con la investigación básica, la cual, a su vez, ha concebido enormes volúmenes de información respecto a la biología molecular humana, y ha definido así un nuevo conjunto de conocimientos bajo el título de la ciencias "ómicas": genómica (origen, evolución, composición y funcionamiento del genoma), proteómica (estructura y función de las proteínas y las modificaciones que un organismo pueda causar), metabolómica (estudio avanzado del conjunto de moléculas como hormonas, moléculas señalizadoras, etc., en una visión dinámica de su interrelación con el individuo), exómica (la ómica relacionada con los exones), etc.[88]. El análisis de las ciencias ómicas y de su importancia en el desarrollo de desenlaces clínicos solo puede ser posible con la integración de equipos que incorporen conocimientos amplios tanto en epidemiologia clínica, bioestadística, biología computacional, bioinformática y minería, como en visualización de datos[89,90], de tal forma que los objetivos de la medicina traslacional se alcancen en plazos prudentes y beneficien rápidamente a la comunidad.

El uso de TIC por parte de los profesionales de la salud debe orientarse a cumplir el objetivo de mejorar sus prácticas mediante un mejor uso de la información, la cual puede ser utilizada tanto en procesos de atención clínica como en el desarrollo de protocolos de investigación. Sin embargo, en la mayoría de los casos los profesionales de la salud no han recibido ningún tipo de entrenamiento formal en informática biomédica, de tal modo que el personal de salud puede suponer, erróneamente, que el uso de computadores en medicina se limita al procesamiento de palabras y textos, al uso de la HCE o a la búsqueda de artículos en bases de datos especializadas[91-93].

La informática biomédica va más allá de la simple instalación de terminales computarizadas en un hospital. Es una ciencia que implica la coalescencia de disciplinas en torno a herramientas computarizadas y digitales que permiten un manejo eficiente de la información en el contexto de la complejidad clínica.

Referencias

1. Van Bemmel JH. The structure of medical informatics: Bibliography on educational courses at the Free University, Amsterdam. Informatics Heal Soc Care. 1984;9:175-80.
2. Shortliffe EH. The science of biomedical computing. Informatics Heal Soc Care. 1984;9:185-93.
3. Winter A. The future of medical informatics. Methods Info Med. 2009;48:62-5.
4. Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med. 2009;48:38-44.
5. García Rojo M. State of the art and trends for digital pathology. Stud Health Technol Inform. 2012;179:15-28.
6. Altman RB. Translational bioinformatics: linking the molecular world to the clinical world. Clin Pharmacol Ther. 2012;91:994-1000.
7. Drew T, Evans K, Võ ML-H, et al. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? Radiogr Rev Publ Radiol Soc North Am Inc. 2013;33:263-74.
8. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363:501-4.
9. Weinfeld JM, Davidson LW, Mohan V. Electronic health records improve the quality of care in underserved populations: a literature review. J Health Care Poor Underserved. 2012;23(Suppl):136-53.
10. Greenes RA, Shortliffe EH. Medical informatics. An emerging academic discipline and institutional priority. JAMA. 1990;263:1114-20.
11. Haux R. Medical informatics: past, present, future. Int J Med Inf. 2010;79:599-610.
12. Friedman CP. A "fundamental theorem" of biomedical informatics. J Am Med Informatics Assoc JAMIA. 2009;16:169-70.

13. Liebovitz D. Meaningful EHR attributes for an era of accountability, transparency, shared decision making, and value assessment. J Leg Med. 2013;34:43-53.
14. Saleem JJ, Flanagan ME, Wilck NR, et al. The next-generation electronic health record: perspectives of key leaders from the US Department of Veterans Affairs. J Am Med Informatics Assoc JAMIA. 2013;20:e175-7.
15. Mantas J, Ammenwerth E, Demiris G, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods Inf Med. 2010;49:105-20.
16. Tachakra S, Wang XH, Istepanian RSH, et al. Mobile e-health: the unwired evolution of telemedicine. Telemed J E-Heal Off J Am Telemed Assoc. 2003;9:247-57.
17. Kuo AMH. Opportunities and challenges of cloud computing to improve health care services. J Med Internet Res. 2011;13:e67.
18. Payne KFB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak. 2012;12:121.
19. Gentles SJ, Lokker C, McKibbon KA. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. J Med Internet Res. 2010;12:e22.
20. Xylinas E, Rouprêt M, Kluth L, et al. Collaborative research networks as a platform for virtual multidisciplinary, international approach to managing difficult clinical cases: an example from the Upper Tract Urothelial Carcinoma Collaboration. Eur Urol. 2012;62:943-5.
21. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA. 2013;309:1351-2.
22. Glaser J, Overhage JM. The role of healthcare IT: becoming a learning organization. J Healthc Financ Manag Assoc. 2013;67:56-64.
23. Bloomrosen M, Detmer DE. Informatics, evidence-based care, and research; implications for national policy: a report of an American Medical Informatics Association health policy conference. J Am Med Informatics Assoc JAMIA. 2010;17:115-23.
24. Corrao S, Arcoraci V, Arnone S, et al. Evidence-Based Knowledge Management: an approach to effectively promote good health-care decision-making in the Information Era. Intern Emerg Med. 2009;4:99-106.
25. De la Iglesia D, García-Remesal M, de la Calle G, et al. The impact of computer science in molecular medicine: enabling high-throughput research. Curr Top Med Chem. 2013;13:526-75.
26. Yang MQ, Athey BD, Arabnia HR, et al. High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. BMC Genomics. 2009;(Suppl 1):I1.
27. Dixon BE, Jones JF, Grannis SJ. Infection preventionists' awareness of and engagement in health information exchange to improve public health surveillance. Am J Infect Control. 2013;41:787-92.
28. Abramson EL, McGinnis S, Edwards A, et al. Electronic health record adoption and health information exchange among hospitals in New York State. J Eval Clin Pract. 2012;18:1156-62.
29. Agrawal A. Medication errors: prevention using information technology systems. Br J Clin Pharmacol. 2009;67:681-6.
30. Galanter WL, Hier DB, Jao C, et al. Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance. Int J Med Inf. 2010;79:332-8.
31. Alexander M. Connect the dots: patient safety and health information technology. J Infus Nurs Off Publ Infus Nurses Soc. 2013;36:85-6.
32. Ancker JS, Kern LM, Abramson E, et al. The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety. J Am Med Informatics Assoc JAMIA. 2012;19:61-5.
33. Khoja S, Durrani H, Scott RE, et al. Conceptual framework for development of comprehensive e-health evaluation tool. Telemed J E-Heal Off J Am Telemed Assoc. 2013;19:48-53.
34. Ehteshami A, Sadoughi F, Saeedbakhsh S, et al. Assessment of Medical Records Module of Health Information System According to ISO 9241-10. Acta Inform Medica Aim J Soc Med Informatics Bosnia HerzegČasopisDruštvaZa Med Inform Bih. 2013;21:36-41.
35. Califf RM, Sanderson I, Miranda ML. The future of cardiovascular clinical research: informatics, clinical investigators, and community engagement. JAMA. 2012;308:1747-8.
36. Sarkar IN. Biomedical informatics and translational medicine. J Transl Med. 2010;8:22.
37. Jing X, Kay S, Marley T, et al. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an EHR prototype based on the Continuity of Care Record standard. J Biomed Inform. 2012;45:82-92.

38. Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the Electronic Health Record for personalized medicine. Hum Mutat. 2011;32:512-6.
39. Bar-Dayan Y, Saed H, Boaz M, et al. Using electronic health records to save money. J Am Med Informatics Assoc Jamia. 2013;20(e1):e17-20.
40. Bui AAT, Taira RK, Goldman D, et al. Effect of an imaging-based streamlined electronic healthcare process on quality and costs. Acad Radiol. 2004;11:13-20.
41. Breitschwerdt R, Iedema R, Robert S, et al. Mobile IT solutions for home health care. Adv Heal Care Manag. 2012;12:171-87.
42. Maglaveras N, Chouvarda I, Koutkias V, et al. Information technology can enhance quality in regional health delivery. Methods Inf Med. 2002;41:393-400.
43. Haynes RB, Wilczynski NL. Computerized Clinical Decision Support System (CCDSS) Systematic Review Team. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review. Implement Sci Is. 2010;5:12.
44. Health Quality Ontario. Robotic-assisted minimally invasive surgery for gynecologic and urologic oncology: an evidence-based analysis. Ont Heal Technol Assess Ser. 2010;10:1-118.
45. McGinness D, Bronnert J, Masarie C, et al. RxNorm's drug interface terminology supports interoperability. J Ahima Am Heal InfManag Assoc. 2013;84:44-7.
46. Rosenbloom ST, Brown SH, Froehling D, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Informatics Assoc Jamia. 2009;16:81-8.
47. Meneghetti A. Challenges and benefits in a mobile medical world: Institutions should create a set of BYOD guidelines that foster mobile device usage. Health Manag Technol. 2013;34:6-7.
48. Kampov-Polevoi J, Hemminger BM. A curricula-based comparison of biomedical and health informatics programs in the USA. J Am Med Informatics Assoc Jamia. 2011;18:195-202.
49. Feng J-Y, Chang Y-T, Chang H-Y, et al. Systematic review of effectiveness of situated e-learning on medical and nursing education. Worldviews Evid-Based Nurs Sigma Theta Tau Int Honor SocNurs. 2013;10:174-83.
50. Muntean V, Calinici T, Tigan S, et al. Language, culture and international exchange of virtual patients. Bmc Med Educ. 2013;13:21.
51. Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. ArtifIntell Med. 2013;57:9-19.
52. Van Belle VMCA, Van Calster B, Timmerman D, et al. A mathematical model for interpretable clinical decision support with applications in gynecology. Plos One. 2012;7:e34312.
53. Lee J. More technology, more risks: top hazards linked to maturing health IT market. Mod Healthc. 2012;42:16-7.
54. Hersh W. Health care information technology: progress and barriers. JAMA. 2004;292:2273-4.
55. Cochran JH. Continuous healing relationships through connectivity. As the nation edges to widespread implementation of EMRs, physicians are concerned about weakening the important doctor-patient relationship. J Healthc Inf Manag Jhim. 2010;24:19-20.
56. Rouf E, Whittle J, Lu N, et al. Computers in the exam room: differences in physician-patient interaction may be due to physician experience. J Gen Intern Med. 2007;22:43-8.
57. Dünnebeil S, Sunyaev A, Blohm I, et al. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inf. 2012;81:746-60.
58. Lin C, Lin I-C, Roan J. Barriers to physicians' adoption of healthcare information technology: an empirical study on multiple hospitals. J Med Syst. 2012;36:1965-77.
59. Sommerhalder K, Abraham A, Zufferey MC, et al. Internet information and medical consultations: experiences from patients' and physicians' perspectives. Patient Educ Couns. 2009;77:266-71.
60. De Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the Internet for pain-related medical information. Patient Educ Couns. 2007;68:86-97.
61. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Heal Inf Manag Ahima Am Heal Inf Manag Assoc. 2012;9:1b.
62. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med. 2013;18:90-2.
63. Guthrie MB. Get real: what will draw physicians to the Web? Physician Exec. 2001;27:36-40.
64. Miller RH, Hillman JM, Given RS. Physician use of IT: results from the Deloitte Research Survey. J Health Inf Manag Jhim. 2004;18:72-80.
65. Bernat JL. Ethical and quality pitfalls in electronic health records. Neurology. 2013;80:1057-61.

66. Nielsen BA, Baum RA, Soares NS. Navigating ethical issues with electronic health records in developmental-behavioral pediatric practice. J Dev Behav Pediatr Jdbp. 2013;34:45-51.
67. Rynning E. Public trust and privacy in shared electronic health records. Eur J Heal Law. 2007;14:105-12.
68. Blobel B. Authorisation and access control for electronic health record systems. Int J Med Inf. 2004;73:251-7.
69. Van Velsen L, Beaujean DJMA, van Gemert-Pijnen JEWC. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak. 2013;13:23.
70. Tuomainen M, Mykkänen J. Interoperability design of personal health information import service. Stud Health Technol Inform. 2012;180:462-6.
71. De la Torre-Díez I, González S, López-Coronado M. EHR systems in the Spanish Public Health National System: the lack of interoperability between primary and specialty care. J Med Syst. 2013;37:9914.
72. Rogoski RR. Bridging the EHR divide. Hospitals, physician practices work toward true interoperability. Healthc Informatics Bus Mag Inf Commun Syst. 2012;29:8-12.
73. Trudel M-C, Paré G, Laflamme J. Health information technology success and the art of being mindful: preliminary insights from a comparative case study analysis. Health Care Manage Rev. 2012;37:31-42.
74. Menachemi N, Matthews MC, Ford EW, et al. The influence of payer mix on electronic health record adoption by physicians. Health Care Manage Rev. 2007;32:111-8.
75. Hebert MA. Impact of IT on health care professionals: changes in work and the productivity paradox. Heal ServManag Res Off J AssocUniv Programs Heal Adm Hsmc Aupha. 1998;11:69-79.
76. Morrison J, Lindberg P. When no one has time: measuring the impact of computerization on health care workers.Aaohn J Off J Am Assoc Occup Heal Nurses. 2008;56:373-8.
77. Stewart MJ, Georgiou A, Hordern A, et al. What do radiology incident reports reveal about in-hospital communication processes and the use of health information technology? Stud Health Technol Inform. 2012;178:213-8.
78. Borycki EM, Kushniruk AW, Bellwood P, et al. Technology-induced errors. The current use of frameworks and models from the biomedical and life sciences literatures. Methods Inf Med. 2012;51:95-103.
79. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res. 2010;10:231.
80. Yan H, Gardner R, Baier R. Beyond the focus group: understanding physicians' barriers to electronic medical records. J Comm J Qual Patient Saf Comm Resour. 2012;38:184-91.
81. Holroyd-Leduc JM, Lorenzetti D, Straus SE, et al. The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence. J Am Med Informatics Assoc Jamia. 2011;18:732-7.
82. Rahadhan P, Poon SK, Land L. Understanding unintended consequences for EMR: a literature review. Stud Health Technol Inform. 2012;178:192-8.
83. Loomis GA, Ries JS, Saywell RM, et al. If electronic medical records are so great, why aren't family physicians using them? J Fam Pract. 2002;51:636-41.
84. Simon SR, McCarthy ML, Kaushal R, et al. Electronic health records: which practices have them, and how are clinicians using them? J Eval Clin Pract. 2008;14:43-7.
85. Hersh W. The health information technology workforce: estimations of demands and a framework for requirements. Appl Clin Informatics. 2010;1:197-212.
86. Tierney MJ, Pageler NM, Kahana M, et al. Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions. Acad Med J Assoc Am Med Coll. 2013;88:748-52.
87. Bornstein SR, Licinio J. Improving the efficacy of translational medicine by optimally integrating health care, academia and industry. Nat Med. 2011;17:1567-9.
88. Bakry R, Rainer M, Huck CW, et al. New stationary phases for enrichment and separation in the "omics" era. Bioanalysis. 2009;1:151-69.
89. Berger B, Peng J, Singh M. Computational solutions for omics data. Nat Rev Genet. 2013;14:333-46.
90. Schneider MV, Orchard S. Omics technologies, data and bioinformatics principles. Methods Mol Biol Clifton Nj. 2011;719:3-30.
91. Sanders CW, Edwards JC, Burdenski TK. A survey of basic technical skills of medical students. Acad Med J Assoc Am Med Coll. 2004;79:873-5.
92. Krause ND, Roulette GD, Papp KK, et al. Assessing medical informatics confidence among 1st and 2nd year medical students. Amia Annu Symp Proc Amia Symp Amia Symp. 2006;989.
93. McGowan JJ, Passiment M, Hoffman HM. Educating medical students as competent users of health information technologies: the MSOP data. Stud Health Technol Inform. 2007;129(Pt 2):1414-8.

Índice analítico*

A

acuerdo
- casi perfecto, 277
- moderado, 277
- pobre, 277
- sustancial, 277
adaptación cultural, 97
adherencia, 32, 91-93, 151, 183, 184, 205, 216, 217, 219, 285, 287, 288, 318, 377, 414, 420, 421, 425, 428-430, 432, 434-439, 465, 466, 470-473, 478, 493, 535, 547, 549, 559
- estrategias, 437
- factores, 435
- - relacionados con el médico y con el equipo de salud, 436
- - relacionados con el paciente, 436
- - relacionados con el sistema de salud, 437
- - relacionados con la enfermedad, 436
- - relacionados con la intervención, 437
- - socioeconómicos, 435
- medición, 439
afijación, 71
- igual, 71
- óptima, 71
- proporcional, 71, 73
afirmación de la consecuencia, 25
agencias, 387
- de documentación, 578
- de evaluación de tecnología, 401, 405, 406
- financiadoras, 302
- gubernamentales, 369, 426, 578
- reguladoras, 376, 426, 432
alcoholismo, 206
aleatorización, 45, 74, 75, 203, 205, 207-213, 216, 217, 221, 222, 239, 307, 310-312, 372, 373, 414, 418, 423, 545, 546, 559, 564, 565, 570
- por bloques, 210
alfa de Cronbach, 109
alzhéimer, 234
amplitud de rango, 102
análisis
- adicional de un subconjunto de datos, 292
- adicional en profundidad, 292
- cualitativo, 459
- cuantitativo, 459
- de costo-consecuencia, 379
- de costo-utilidad, 379
- de decisiones, 325
- - bivariado, 240
- - clasificación de los problemas de decisión, 328
- - de costo-beneficio, 353, 360, 378
- - de costo-efectividad, 355, 356, 360, 379, 383, 393

* *v.* = véase; *v. t.* = véase también.

- - de costo-utilidad, 360, 379
- - de decisiones, 36, 183, 325, 327, 329, 330, 359, 376, 380, 403, 533
- - de minimización de costos, 354, 356
- - de regresión, 218
- - de sensibilidad, 195, 263, 305, 312, 316, 377, 380, 381, 383, 403
- - económico, 170, 183, 376-378, 380, 402, 403, 405, 543
- - exploratorio, 456, 461, 463, 502, 503
- - factorial, 102
- - fases, 326
- - - elección, 326
- - - independencia
- - - linealidad, 327
- - - orden completo, 327
- - - simetría, 327
- - final, 182, 416, 456, 460-463
- - interino, 218, 414, 459
- - multivariado, 193, 458, 459, 523
- - plan de, 9, 148, 150, 153, 163, 191, 264, 266, 378, 417, 454-463
- - por intención de tratar, 205, 217, 545, 546
- - por protocolo, 217, 546
- - univariado, 461, 520
- de sensibilidad, 380
- de sobrevida, 400
- estratificado, 242
- exploratorio, 463
- factorial, 102
- interino, 459
- multivariados, 458
- por intención de tratar, 218, 546
- por protocolo, 217, 546
- secuencial, 202
- secundario, 292
- simple, 462
- ***v. t.*** **criterios de decisión, representación**
anemia, 95, 219, 303, 306
- ferropénica, 307
angina, 132
- inestable (AI), 381
anomalía congénita, 424
ansiedad, 281
antepropuesta, 147
años de vida
- ajustados por calidad, 354, 359, 367, 379, 403
- esperados estandarizados, 31, 35, 36, 78, 160, 181, 300, 352, 421, 443, 459, 460, 541, 580
- esperados por cohorte, 38, 59, 68, 69, 74, 77, 85, 115, 116, 118, 121, 127, 179, 211, 213, 230-232, 235, 236, 244, 245, 247-267, 356, 358, 374, 375, 462, 542
- potencial, perdidos, 263, 360, 547
- saludable, 354, 355, 360, 367
- saludables ajustados por discapacidad, 354, 360, 367, 379, 397, 403, 548, 566
apareamiento, 227, 240, 245, 260
- efectividad de, 4, 32, 118, 121, 127, 172, 184, 201, 202, 213, 217, 219, 245, 300, 301, 306, 311, 314, 317, 334, 351, 354, 355, 357, 376, 378, 380-382, 390, 392, 393, 397, 398, 401-403, 434, 438, 529, 547, 549, 564, 566
- ***v. t.*** **emparejamiento**
apendicitis, 4, 541
- aguda, 4
- - perforada, 542
aplicabilidad de los resultados, 7, 98, 151, 164, 169, 207, 214, 217, 219, 302, 335, 338, 340, 342, 412, 540, 553, 561
apreciación crítica
- de la literatura, 182, 211, 338, 399, 532, 534, 538, 540, 544, 545, 549-551
- - evaluación
- - - generalizabilidad, 549
- - - importancia de los resultados, 547
- - - validez interna, 545
- - pregunta, 541
- - tipo de estudio, 542
- cualitativa, 295
- - método, 294
- - muestreo, 289
- - pasos, 287
- - programas informáticos, 293
- - reporte, 293, 294
- - sesgos, hipótesis y supuestos, 294
- - técnicas, 282
- - - entrevistas, 284
- - - grupos focales, 285
- - - observación, 282
- - validez y confiabilidad, 288
apuesta estándar, 360, 452
árbol de decisiones, 330
- *v. t.* **toma de decisiones**
archivos
- electrónicos, 375, 431
- de trascripción, 466
- planos, 12, 38, 293, 375, 427, 429, 431, 466, 482, 483, 492 499, 500
áreas bajo la curva, 86
argumentos, 16
- deductivos, 16
- lógicos, 15
arritmias, 375
- cardíacas, 203
- postinfarto, 434

- supraventriculares, 372
artículos
- agregativos, 36
- originales, 36
artritis, 436
- gotosa aguda, 79
- reumatoide, 201, 206, 209
asfixia neonatal, 306
asignación, 68
- al azar, 545
- aleatoria, 166, 209, 210, 212, 213, 219, 221, 222, 244, 418, 423, 545, 546, 558, 564, 565
- ciega, 545, 546
- enmascarada, 211, 546
- observacional, 245
asma, 201, 316, 424
- bronquial, 550, 579
asociación
- calculadoras, 126
- causal, 35, 85, 236, 245, 247, 257, 266, 372
- causalidad, 19, 20, 245
- colaborativa, 555
- confusión de la, 260, 265, 458, 462, 536
- "cruda", 56, 462
- de variables, 21, 113, 269
- entrevistas, 185
- evaluación, 213, 239, 241, 248, 249, 257, 266, 458, 462
- exposición y desenlace, 55-57, 67, 68, 84, 85, 116, 117, 119, 121, 218, 227, 228, 230, 234, 236, 239, 241, 245, 265, 456-458, 519, 520
- exposición y riesgo, 121, 178, 192, 250
- frecuencia, 19, 55, 59, 113, 121, 123-125, 176, 194, 208, 264
- hipótesis de, 67, 244, 260, 264, 266
- incidencia, 248
- independiente, 57
- interacción, 462
- intervención y desenlace, 217, 218, 424
- medida de, 113, 116, 118, 120, 122, 126, 190, 219, 238, 241, 259, 262, 264, 269, 279, 280, 312, 371, 458
- - razón de riesgo, 190
- - riesgo relativo, 190, 256, 314, 315, 373
- modelado de, 75, 86
- prevalencia, 122
- probabilidad, 20
- real, 256, 257, 536
- tamaño de la, 79
ataque cerebrovascular (ACV), 338, 353, 381, 382
- hemorrágico, 353
- isquémico, 353
- isquémicos transitorios (AIT), 215
- vasculares cerebrales, 215
atopía, 18
atributos, 245
- identificadores, 469
ausencia de sesgo, 255
autismo infantil, 541
autoadministrados o administrados por terceros, 445
autonomía, 189, 221, 294, 332, 394, 395, 553, 568
- disminuida, 553
- respeto por la, 221, 394
autores fantasma, 576
autoría, 40, 413, 570, 572-587
- agradecimientos y reconocimientos, 580
- autorregulación y control, 585
- compromiso, 583
- criterios de autoría, 583
- estudios sobre autoría, 581
- justificada, 572
- méritos, 40, 413, 570, 572-587
- otras soluciones, 584
- problemas, 573
- propuestas de soluciones, 582
- publicaciones en internet, 585
- recomendaciones, 583
- ***v. t.*** **problemas de la autoría**
autoselección de los participantes, 258
azar, 5, 6, 59, 60, 62, 63, 68, 73, 78, 114, 116, 117, 123, 125, 126, 137, 171-176, 199, 200, 205, 208, 209, 211-213, 216, 220, 222, 231, 237, 239, 245, 268-274, 307, 313, 314, 318, 375, 418, 535, 536, 543, 545, 548, 564-566

B

base
- de estudio primaria, 235
- de datos, 30, 31, 35, 37-39, 153, 205, 230, 307, 308, 338, 417, 420, 455, 461, 465-467, 470, 473, 481-483, 485, 488, 490, 492-494, 500, 503
- - tamaño de muestra, 63, 68, 205, 252, 573
- empírica, 6
- secundaria, 375
- ***v. t.*** **construcción de bases**
beneficios, 33, 182, 333, 351, 354, 357, 361, 390-392, 394, 406, 554, 558, 579
- balance, 182, 333, 394, 406, 556
- costo, 352-354, 360, 378
- de la duda, 320
- diferencia, 354, 357

- mayor, 351, 370, 393, 554, 565
- posible, 169, 170
- resultados, 558
- riesgo, 33, 368, 405
- tecnología, 393, 598
biblioteca Cochrane, 38, 39, 307, 406, 535
bioética, 552, 563, 568
bondad de ajuste, 26
bronconeumonía, 275
bronquiolitis, 316, 541, 550
bronquitis crónica, 206, 260
búsqueda
- de casos, 131
- de información, 36, 37, 147, 337, 398, 483, 533, 537
- de la literatura, 30, 31, 33, 35, 37, 39, 41, 148, 301, 304, 308, 309, 534, 579, 584
- de los estudios, 304, 305, 307, 308, 310
- especificidad, 31, 86, 136-138, 143
- - medición, 135
- secundaria, 308
- sensibilidad, 30-33, 35, 37, 48, 81, 86, 96, 98
- - medición, 134

C

cálculo
- de las suertes, 141
- de magnitud, 14, 302, 549, 557
- del costo, 353, 361, 381, 481
cálculos en la vesícula, 131, 180
calidad de vida, 10, 19, 23, 91, 95, 96, 102, 215, 281, 337, 350, 358-360, 378, 379, 390, 391, 397, 434, 443, 444
calificación
- de la información, 48, 101, 183, 320, 391, 401, 476-478
- moderada, 339
campos de tipo numérico, 269, 275, 487
cáncer, 206, 226, 275, 369, 424, 436, 444, 529, 536
- de colon, 123
- de cuello uterino, 131
- de esófago, 230
- de laringe, 360
- de páncreas (CAP), 55, 56, 184
- de pulmón, 6, 55, 56, 169, 176, 177, 184, 227, 228, 241, 266, 536
- de seno, 178, 227
- de vejiga, 59, 178, 184, 258, 259
- gástrico, 262
- hepático, 368
- prostático, 131
- vaginal, 227
capital humano, 349, 350, 361, 363
captura, 156, 250, 291, 355, 420, 421, 455, 472-476, 478, 481, 482, 589
características operativas, 86, 129, 132-134, 137, 138, 143, 166, 171, 180, 181, 268, 537, 580
caras de Chernoff, 523, 525-527
carcinoma de colon, 372
- de páncreas, 237
- de pulmón, 227, 273
- de vejiga, 59
cardinalidad, 470
caries, 169
casos y controles, 35, 36, 68, 77, 78, 84, 85, 121, 168, 171, 174-178, 186, 193, 196, 226-243, 250, 266, 301, 371, 372, 374, 375, 462, 532, 542
- cruzados, 375
- - elección, 237
- - fuentes, 237
- - - grupos especiales, 238
- - - población general, 237
- - - sujetos hospitalizados o en tratamiento, 237
- - principios de comparabilidad y selección, 234
- - - de control de la confusión, 236
- - - de eficiencia, 236
- - - de la base del estudio, 235
- - - de precisión comparable, 236
- historia, 226
- incidentes, 230
- medida de asociación y análisis, 241
- prevalentes, 230
- razón de controles a casos, 239
- - emparejamiento, 240
- tipos de casos, 230
- - anidados en una cohorte, 231
- - basados en casos, 230
- - caso-cohorte, 231
- - cruzado de casos, 232
- ventajas y desventajas, 240
- ***v. t.*** **definición y selección de casos y controles,** tipos de sesgos
cataratas congénitas, 227
categorización, 49, 51, 239, 273, 275, 293
causalidad, 19, 20, 23, 118, 169, 175, 178, 179, 186, 189, 193, 194, 226, 245, 247, 256, 266, 339, 541, 542
cefalea, 304
censo, 63, 212, 463
certeza diagnóstica, 180
choque cardiogénico, 172
ciego, 173, 205, 418, 566
cierre de un proyecto, 431
- equipos e insumos, 431

- información, 431
- informes, 432
cirrosis, 230, 368
Cochrane, 38, 39, 307, 406, 535
codificación de datos, 52, 460, 590
código de Nuremberg, 552, 553, 562
coeficiente
- de concordancia, 110
- de correlación, 110, 276, 277
- - de concordancia (CCC), 277
- - de Pearson, 108, 110-112, 276, 277, 279, 280
- - intraclase (CCI) de Lin, 277
- - r, 277
- de la regresión logística, 110, 276, 277
cohortes, 34, 36, 68, 78, 85, 167, 171, 174, 178-180, 186, 196, 226-228, 232, 241, 247, 248, 256, 257, 266, 301, 302, 358, 372, 374, 398, 400, 462
- ambidireccional, 250
- analítica, 38, 59, 68, 69, 74, 77, 85, 115, 116, 118, 121, 127, 179, 211, 213, 230-232, 235, 236, 244, 245, 247-267, 356, 358, 374, 375, 462, 542
- - de múltiples cohortes, 248
- - de una cohorte, 248
- descriptiva, 247
- - asociación y causalidad, 245
- - etapas del diseño y la conducción de un estudio analítico de cohorte, 254
- - - análisis, 264
- - - asignación, 256
- - - ensamblaje, 254
- - - muestreo, 255
- - - seguimiento, 260
- - - selección, 254
- - proporción y tasa de incidencia, 252
- - sesgo de detección temprana (***lead-time bias***), 253
- - tiempo, 248
- - - calendario, 250
- - - causal, 248
- - - de incepción, seguimiento, incidencia y ensamblaje de la cohorte, 251
- concurrente, 250
- de incepción, 251
- de origen, 251
- histórica, 250
cointervención, 214, 547
colaboración Cochrane, 38, 39, 300, 302, 304, 306, 311, 314, 317, 338, 398, 400
colesterol sérico, 443
colon irritable, 108, 201
comité de revisión, 218
- externos, 414
comoderador, 287
comparabilidad de grupos, 96, 176, 195, 199, 214, 217, 221, 230, 234, 236, 237, 373, 379, 403, 418, 546
comparador, 244, 378, 399
compensación temporal, 359
competencia de prioridades, 435
complicaciones
- en la retina, 183
- renales, 183
comportamientos, 92
comunalidad, 106
concordancia, 12, 57, 108, 110, 171, 180-182, 186, 268-273, 275-279, 310, 312, 420, 435, 437
- interobservador, 269
- prueba de, 110, 182
conducción, 17, 115, 148, 149, 151, 153-155, 160, 206, 211, 218, 221, 231, 244, 254, 255, 258, 265, 266, 286, 300, 302, 311, 317, 340, 387, 402, 411-433, 436, 438-440, 446, 455, 563, 570, 572, 576, 582, 584
- de estudios clínicos, 411
- - consideraciones previas, 415
- - - aprobación del estudio, 416
- - - evaluación de factibilidad, 415
- - - planeación, 416
- - desarrollo del estudio, 429
- - - estrategias, 429
- - finalización, 430
- - grupo de trabajo, 412
- - - comités externos, 414
- - - investigadores, 412
- - multicéntricos, 411
- - multisitio, 411
- - planeación, 412
- - tendencias actuales, 432
- - un solo sitio, 411
- ***v. t.*** **cierre de un proyecto, inicio del estudio,** monitorización, recolección de información
confiabilidad, 16, 58, 76, 78, 80, 81, 87, 96-98, 101, 109, 110, 205, 269, 288, 289, 294, 295, 337, 427, 450, 455, 460, 544
confianza, 57, 457
confirmación clínica, 131, 132, 263, 484, 495
conformidad, 269
confusión, 371, 458
conglomerado, 72
conjuntivitis crónica, 237
conocimiento, 12, 45
- científico, 11
- - argumentación lógica, 15

- - definiciones, 12
- - lenguaje, 14
- - validez, 16
- - veracidad, 16
- marcadamente particular, 13
consenso, 36, 185, 286, 295, 306, 310, 311, 320, 331-335, 341, 352, 380, 381, 530, 534, 535, 557, 560, 563, 565
- con base en evidencia, 535
- formal, 332, 333, 341
- informal, 332, 333
consentimiento informado, 153, 165, 207, 221, 293, 294, 395, 414, 422-424, 426, 429, 553, 554, 556-558
consistencia, 26, 58, 109, 182, 269, 289, 303, 311, 317, 400, 428, 429, 460
- interna, 109
construcción
- de bases, 481
- - definiciones del modelo de datos, 482
- - - entidades, 482
- - - valores permitidos, 458
- de novo, 96, 97
- de preguntas, 15, 45, 96, 151, 159, 161, 231, 291, 293, 327, 358, 446, 465, 466, 473, 481-483, 485, 487, 489, 491-494, 505, 506, 508-510, 512, 518, 590
constructo, 91-97, 101, 102, 108, 109, 248, 291
- definición de, 93
contaminación, 26, 202, 261, 350
contenido empírico, 25
contraindicaciones, 372
control, 7, 27, 73, 81, 152, 155, 167, 171, 216, 231, 234, 241, 282, 286, 336, 388, 395, 431, 434
- adecuado, 122, 124, 155, 159, 183, 217, 219, 233, 237, 261, 287, 301, 307, 309, 361, 437, 439
- análisis, 264, 326
- ausencia de, 68
- autocontrol, 374, 375, 585
- calidad, 191, 456, 460, 585
- confusión, 340, 569
- de errores, 32
- de las enfermedades, 437
- de síntomas, 530
- definición, 242
- difícil, 168
- estrategias, 245, 315, 389, 432
- estudios, 411
- exitoso, 203
- exposición, 23, 67, 117, 199, 244, 248, 304
- externo, 108
- factor, 68, 176, 435
- grupo, 83, 117, 119, 152, 171-173, 175, 176, 200, 201, 208, 209, 211, 213, 214, 216, 217, 220, 221, 226, 234, 238, 239, 256, 257, 321, 372, 417, 531, 542, 545, 546, 548, 550, 563, 564, 566
- inadecuado, 122, 437, 569
- intervención, 65, 76, 172, 198, 203, 204, 209, 211, 220, 304, 534, 559
- mecanismos, 170, 209, 427
- periodo, 232
- población, 32, 232
- principio, 235, 236
- procedimientos, 420, 427
- proceso, 209, 236, 261
- programas, 388
- recolección, 466, 471
- respuesta de, 451
- seguimiento, 156, 159, 231, 413
- selección, 237, 238
- sesgos, 371
- ***software*, 156**
- suficiente, 6
- sujeto, 236, 237
- tratamiento, 211, 418, 546, 566
- variabilidad, 72
- variables, 53, 56, 57, 87, 171, 193, 218, 219, 328, 416
conversión de datos, 30, 142, 143, 361, 444
coordinador del estudio, 416, 419, 421, 423, 428, 431
copia de seguridad, 418, 490, 544, 585
correlación, 87, 103-105, 107-111, 171, 180-182, 186, 268, 269, 271, 273, 275-277, 279, 280, 310, 581
- lineal, 279
- - de Pearson, 108, 110, 111
- - de Spearman, 108, 110
- medidas de concordancia para variables numéricas, 275
- - diferencia entre la correlación y la concordancia, 277
- y concordancia, 268
- ***v. t.* métodos para evaluación de concordancia**
corte transversal, 188
- análisis, 190
- - limitaciones, 194
- - ventajas, 193
- diseño, 189
- - consideraciones éticas, 189
- - fuentes de datos, 190
- - población blanco, 189
- - - instrumentos y técnicas de medición, 190

- - - población total o muestra, 189
- - - representatividad de la muestra, 189
- - - selección de variables y escalas de medición, 190
- - pregunta, 189
- reporte y publicación, 194
costo, 19, 23, 40, 157, 159, 170, 179, 180, 183, 186, 201, 228, 234, 237, 242, 307, 326, 332, 350, 351-365, 368, 370, 376-379, 381-383, 389, 391, 393, 398, 402, 403, 419, 431, 437-439, 481, 531
- anual equivalente, 115, 381-383
- beneficio, 378
- de capital, 311, 363
- de oportunidad, 350, 351, 361, 376
- directos, 23, 362
- - médicos, 362
- - no médicos, 362
- evitables, 363
- fijos y variables, 362
- futuros, 431
- incremental, 356, 378, 379, 381, 383
- indirectos, 157, 362, 363
- inducidos, 356
- intangibles, 363
- neto, 361, 378, 379
- no evitables, 363
- tangibles, 363
creencias populares, 436
criterios, 24, 64, 66, 67, 74, 148, 152, 164, 191, 193, 195, 204-208, 230, 239, 257, 290, 305, 306, 340, 415, 416, 421, 423, 427, 456, 461, 532, 536, 549, 563
- de admisión, 74, 148, 207, 208, 415, 421, 422, 532
- de elegibilidad, 195, 205-207, 305, 310, 340, 423, 460, 556
- de exclusión, 152, 206
- de decisión, 329
- - incertidumbre, 329
- - - media con varianza acotada, 329
- - - valor esperado, 329
- - - varianza mínima y media acotada, 329
- de inclusión, 24, 64, 66, 67, 74, 152, 164, 193, 204-206, 230, 239, 257, 290, 305, 416, 423, 427, 456, 461, 536, 549, 563
- diagnósticos, 191, 195, 230, 306, 423
- incertidumbre total, 329
- - Laplace, 326
- - maximax, 326
- - maximin, o criterio de Wald, 329
- problema con certidumbre, 329
cuartiles, 515
cuchilla de Ockham, 27
cuerpo de evidencia, 400
cuestionarios, 96, 98, 100, 152, 184, 190, 214, 216, 228, 229, 232, 233, 237, 404, 416, 443-447, 449, 451-453, 459, 473, 481, 482, 557, 582
- planeación, 445
- - información, 445
- - preguntas, 445-447, 449
- - - abiertas, 449
- - - cerradas, 450-452
curiosidad, 17, 21, 532
- ***v. t.*** **pregunta**
curso clínico, 19, 23, 27, 246, 251, 253, 388

D

datos estadísticos
- alfanuméricos, 493
- análisis, 503
- antecedentes, 25
- atípicos extremos, 515
- atípicos remotos, 515
- continuos, 108, 110, 181, 314
- discretos, 47, 48, 283, 358, 508
- dos variables, 519
- - gráfico, 520
- - - pirámides poblacionales, 521
- - tablas para dos variables, 519
- dos variables cuantitativas, 523
- - caras de Chernoff, 523
- - estrella, 525
- - gráfico, 523
- - red, 525
- gráficos, 506
- - partes, 506
- información, 503
- tablas y gráficos, 505
- - de frecuencias, 505
- - partes, 505
- variable cualitativa, 507
- - gráficos, 508
- - - barras simples, 508
- - - pictograma, 508
- - - torta, o pie, 509
- - tablas, 507
- variable cuantitativa continua, 515
- - gráficos, 517
- - - histograma, 517
- - - línea, 519
- - - polígono, 517
- - - tallo y hojas, 517
- - tabla de frecuencias, 515

- variable cuantitativa discreta, 510
- - gráficos, 512
- - - ***box plot*** (caja y bigote), 513
- - - caja de letras y valores, 515
- - - punto para una variable, 513
- - tablas, 510, 511
- *v. t.* **construcción de bases, modelado,** preparación de datos, presentación estadística de datos
decisiones, ***v.*** toma de decisiones
declaración de Helsinki, 153, 165, 189, 416, 553, 554, 560, 562, 568
definición, 403
- de caso, 174, 185, 230-232, 235, 374, 375, 414, 420, 427
- operativa, 403
demencia
- frontotemporal, 233, 237
- vascular, 233
densidad de incidencia, 115, 116, 231, 262, 375
depresión, 96, 233, 281
depuración, 134, 156
- de la base de datos 308, 417, 455, 465, 467, 481-483, 485, 493, 495, 503
- limpieza y, 456
dermatitis atópica, 18
desarrollo de novo, 28, 97
desenlace, 23, 52, 65, 67, 68, 126, 399, 534
- adverso, 550
- aparición, 117, 250, 252
- asociación, 265
- búsqueda, 375
- categorías, 339
- causa de, 245
- clasificación, 258, 259
- comparabilidad, 217
- control, 173, 232
- criterios, 542
- de interés, 53, 67, 74, 78, 79, 82, 85, 96, 121-123, 177-179, 219, 226, 228-230, 232, 234-237, 247, 255, 262, 264, 319, 371, 375, 400, 546-548
- - en seguridad, 547
- - indeseable, 117
- desfavorable, 121
- discreto dicotómico, 265
- dicotómico, 77, 84, 85, 203, 218, 265
- efectividad, 354, 358
- - indirecta, 393
- en estudio, 237, 241
- enfermedad, 246, 247, 254
- escogido, 108, 178
- específico, 23, 24, 96, 247, 359
- evaluación de, 68, 86, 306, 317-319, 322, 379, 401, 414
- exitoso, 203
- exposición y, 35, 55, 56, 117, 121, 176, 179, 219, 245, 249, 250, 261, 262, 340, 373, 519, 541
- frecuencia, 57, 67, 116, 119, 121, 123, 174, 178, 199, 208, 217, 244, 247, 261, 262, 265, 266, 419, 548-550
- futuro, 108, 116, 249
- incidencia, 65, 68, 259, 260, 262
- indeseable, 116, 117, 550
- mala calificación, 235
- medida de, 166
- no desenlace, 228
- no expuesto, 249
- poco común, 34
- positivo, 215
- predicción, 96, 171, 231
- preguntas, 20, 23
- previsible, 199
- principal, 204, 205, 215, 248, 251
- probabilidad, 59, 88, 211, 212
- pronóstico, 372
- prospectivo, 250
- punto de, 124
- regla de oro, 319
- retrospectivo, 250
- riesgo, 121, 207, 241, 542, 549
- seguimiento, 260
- sustituto, 550
- variabilidad, 69, 74
- variables, 53, 55, 86, 152, 171, 189, 199, 204, 215, 244, 264, 265, 372, 457, 569
- - dependiente, 198, 200
definición y selección de casos y controles, 232
- grupos especiales, 234
- hospitalarios, 233
- poblacionales o de la comunidad, 234
desviación
- estándar, 62, 71, 82, 83, 276, 277, 512
- sistemática, 59, 130, 188
diabetes, 27, 57, 65, 131, 137, 139, 170, 180, 206, 228, 229, 261, 264, 382, 383, 436-438, 509, 559
- con infarto cerebral, 238
- gestacional, 238, 306
- ***mellitus***, 137, 261, 505
- - tipo 1, 27, 57
- - tipo 2, 27, 57, 73, 88, 381
diagnóstico, 132
- descarte de, 132
diagramas
- de Gantt, 155

- de torta o ***pie***, 508, 509, 511
- de entidad-relación, 466
- de tallo y hoja, 510
diálisis crónica, 211
diarios personales, 285
diarrea, 550
diccionario de términos MeSH, 33
diferencia mínima, 79, 458
digitación, proceso de, 417
dirección
- de la asociación, 256
- del efecto, 403
Disability Adjusted Life Years (DALY), 360, 403
discapacidad, ***v.*** años de vida
discusión, 19, 24, 26, 185, 193, 196, 220, 255, 258, 265, 272, 285-288, 310, 312, 332, 333, 341, 373, 379, 404, 405, 419, 428, 474, 475, 567, 569
diseño, 9, 17, 24, 34, 62, 68, 72, 73, 77, 80, 83, 88, 89, 121, 154, 158, 160, 161, 163, 164, 189, 202, 204, 208, 218, 219, 221, 227, 229, 232, 234, 235, 245, 265, 276, 282, 285, 295, 299-301, 331, 336, 375, 378, 379, 398, 400, 412, 416, 417, 445, 456, 461, 484, 495, 543, 554-556, 569, 581, 590, 593
- adecuado, 149, 170, 421, 566
- aplicabilidad, 98
- apropiado, 168, 169, 183, 186
- calidad, 240, 305, 307, 313, 339, 564
- características, 339
- causalidad, 178
- - cohortes, 178, 179, 186, 231, 247, 248, 254, 256, 266, 462
- concurrente, 152
- histórico, 152
- metodológico, 34
- mixto, 152, 295
- muestral, 72
- - bietápico, 80
- - polietápico, 72
- multicéntrico, 411
- multisitio, 411
- un solo sitio, 411
- coordinador, 432
- - equipo, 464, 572, 583, 584
- corte transversal, 189, 193, 194
- costo-beneficio, 360, 361
- de estudio, 8, 32, 33, 98, 195, 198, 204, 205, 260, 293, 310, 311, 367, 411, 412, 429
- - casos y controles, 176, 227-229, 231, 462
- - - desventajas, 240, 241
- - - ventajas, 240
- - integrativo, 170, 300
- de los modelos longitudinales de análisis, 261
- de panel, 261
- del estudio, 8
- descontinuación, 203
- - limitaciones, 203, 204, 318, 320, 321, 401
- descriptivos, 174
- - reporte, 174
- distribución simétrica, 209
- efecto, 80, 82
- efectos, 260
- error sistemático, 322
- especificidad, 180
- estadístico, 156, 375
- estrategia, 78, 150, 299, 455, 457, 461, 463
- etapas, 255, 304, 593
- evaluación, 232, 288, 318, 373, 402
- - del riesgo, 373
- experimental, 213
- experimento cruzado, 200, 201, 375
- experimento, 170
- - clínico, 170, 171, 204
- flexibilidad, 283, 288, 290
- frecuencia, 261
- híbrido, 232
- histórico, 152
- ideal, 168, 186, 542
- de investigación, 17, 24, 83, 88, 148, 149, 151, 152, 161, 166-170, 226, 276, 282, 285, 303, 304, 336, 554, 581
- - farmacoepidemiología, 373, 374
- - - analíticos, 373
- - - descriptivos, 373
- - formal, 182
- - secundaria, 170
- metaanálisis, 182
- metodología, 189, 567
- - ética, 189
- mixto, 152
- muestra, 458
- muestral polietápico, 72
- no apropiado, 168, 169
- otras técnicas, 185
- - clases de preguntas, 169
- - relación pregunta-diseño, 169
- - selección, 168
- panel, 261
- patrones, 470, 476, 478
- periodo de lavado, 565
- placebo, 555
- plan de análisis, 461, 462, 464
- por grupos, 202
- pregunta, 169, 170, 204, 304, 307, 312, 456
- principios, 200, 234

- protocolo, 397, 425, 556
- seleccionado, 152, 169
- sensibilidad, 180
- sesgo, 59, 179, 238, 266, 301, 303, 317, 340, 372, 373, 375, 537
- validez, 555
- variables, 57
- ventajas, 173
- ***v. t.*** **estudios integrativos, investigación,** métodos cualitativos, modalidades según el sujeto de estudio, recolección de datos
disfunción eréctil, 114
dislipidemia, 27, 73, 170, 204, 206, 208
disnea, 281
- de ejercicio, 137
- de esfuerzo, 530
disposición a pagar, 361, 393
distribución
- de frecuencias absolutas, 508, 511-513, 516, 517
- de frecuencias relativas, 509
- normal, 82, 83, 87, 461, 512
- percentilar, 51
doble
- enmascaramiento, 566
- digitación, 461
- simulación, 216
documentos fuente, 414, 427-429, 432
dolor
- abdominal agudo, 541
- neuropático por diabetes, 172
drogadicción, 206
dualismo, 11

E

Ébola, 168
eclampsia, 117-121, 124-126, 204
economía, 389
edema pulmonar, 172, 281
efectividad, 4, 19, 20, 23, 27, 32, 34, 39, 118, 121, 127, 169, 170, 172, 183, 184, 186, 201-203, 213, 214, 217, 219, 245, 300, 301, 311, 314, 317, 325, 332, 334, 349-358, 360, 367-370, 376-383, 386, 389-393, 396-399, 401-403, 405, 434, 437-439, 529, 533, 547, 549, 559, 563, 564, 566, 591
- del apareamiento, 227, 240, 245, 260
- incremental, 356
efecto(s), 36, 199, 200, 215, 216, 248, 304, 305, 313, 314, 424, 547
- adversos, 36, 304, 424
- aleatorios, 305, 313, 314
- de cohorte, 248
- de período en estudios transversales, 83, 188, 193, 196
- fijos, 305, 313, 314
- Hawthorne, 199, 200
- placebo, 200, 215, 216, 547
- precisión de la estimación, 403
eficacia, 19, 20, 23, 172, 183, 199, 204, 214, 215, 217, 218, 350, 352, 355, 367, 369, 377, 380, 391, 392, 412, 565, 566
eficiencia, 19, 31, 34, 71, 72, 129, 207, 217, 227, 235, 240, 350, 353, 355-358, 361, 374, 376, 377, 393, 530, 551, 573, 574
- distributiva, 356
- estadística, 236
- operacional, 237
- productiva, 355, 356
El camino a la realidad, **45**
elección racional, 325
- alternativas o acciones, 326
- consecuencias, 326
- decisor racional, 326
- elementos, 326
- estados de la naturaleza, 326
- regla de decisión, 326
emparejamiento, 57, 227, 231, 240, 265, 375
encefalitis
- por herpes, 129
- rábica, 16
encuestas, 73, 114, 158, 166, 188, 190, 191, 193, 332, 394, 398, 404, 444, 446, 459, 540, 557
- de prevalencia, 188
enfermedad(es)
- acidopéptica por reflujo, 132
- cardiovascular isquémica, 180, 266, 381, 579
- cerebral, 206
- cerebrovascular, 266
- coronaria, 61, 178, 179, 206, 529, 580
- crónica, 209, 227
- dermatológicas crónicas, 201
- diverticular no complicada, 237
- gastrointestinales, 436
- hepática, 206
- - por alcohol, 230
- infecciosa grave, 357
- muscular, 206
- periodontal severa, 557
- prevalentes, 226
- psiquiátricas, 206
- pulmonar, 239, 436, 456
- pulmonar obstructiva crónica (EPOC), 34, 266
- raras, 227, 242
- renal, 65

- - avanzada, 206
- tracto respiratorio, 269
- trasmisión sexual, 184
- tromboembólica, 529
- tromboembolismo pulmonar (TEP), 142
- vascular, 206
- vía biliar, 529
enfoque de la medición, 48
- cualitativo, 48
- cuantitativo, 49
- escala, 49
- - intervalo, 52
- - nominal, 50
- - ordinal, 50
- - razón, 52
- - variable continua, 52
- - variable discreta, 49
- - - operaciones, 51, 53
- explicativo, 18
engaño justificado, 559
enmascarados, 215, 216, 239, 541, 542, 546, 566
enmascaramiento
- apertura, 172, 173, 211, 214, 311, 318, 414, 418, 531, 546, 565, 566
enmiendas al protocolo, 413
ensamblaje de novo, 374
enterocolitis
- necrotizante, 320
- severa, 321
entrevista, 93, 114, 185, 190, 231, 237, 259, 284, 285, 288, 289, 416, 422, 445, 470, 475, 477
- conjunta, 285
- en profundidad, 259, 284
- estructurada, 185, 259, 284
- no estructurada, 283
- semiestructurada, 284, 288
epidemias, 226
epilepsia, 172
epitelioma de células escamosas del labio, 227
EPOC, 447
equamax, 105
equilibrio clínico, 564
equipo
- de trabajo ampliado, 335
- técnico, 335
equivalencia translingüística, 99
error, 25, 69, 80, 81, 83, 84, 86, 108, 202, 218, 219, 255, 257, 258, 260, 265, 304, 315, 543, 548
- α, 458
- β, 236, 458
- aleatorio, 25, 45, 59, 255, 263, 264, 543, 548
- azar, 208
- común, 37, 444, 467
- control, 171, 486
- corrección, 461, 503
- de decisión, 239
- de muestreo, 78
- diagnóstico, 233
- diferenciales, 236
- ensayo, 5
- estándar, 124, 125, 271, 274, 275, 312
- fuentes de, 17, 58, 169, 182, 265, 304, 543, 544, 547
- - de variabilidad, 57
- interpretación, 241
- límites, 25
- medición, 45, 57, 58, 100, 110, 263, 276
- muestral, 63, 70, 76, 78, 80
- - grande, 63
- porcentaje, 239
- precisión comparable, 236
- probabilístico, 69, 70, 72, 73, 256, 381, 383, 403
- representación gráfica, 508, 509
- sistemático, 219, 257, 258, 265, 304
- tipo I, 69, 80, 81, 83, 84, 86, 202, 208, 218, 458
- tipo II, 69, 81, 86, 108, 208, 458
- tipo III, 21
- - sistemático, 45, 59, 130, 219, 257, 265, 304
- *v. t.* tipos de error
escalas, 49-54, 95-104, 106-111, 152, 165, 188, 190, 206, 229, 264, 265, 269, 271-273, 277, 311, 334, 359, 389, 444, 451, 452, 459, 460, 482, 502, 513, 543, 558
- búsqueda y selección, 98
- de calidad de vida relacionadas con salud, 358
- de intervalo, 52, 53, 359, 459
- de razón, 52, 459
- de rango, 49, 51, 97
- instrumentos, 95
- - construcción y validación de novo, 96
- - discriminativos, 96
- - evaluativos, 96
- - predictivos, 96
- medición, 49
- nominal, 50
- ordinal, 50, 265, 269, 273, 359
- proceso, 97
- traducción, 99
- - sentido inverso, 99
- prueba piloto, 99-101
- sensibilidad al cambio, o capacidad de respuesta, 110, 111
- *v. t.* evaluación de escalas, validación de escalas
esofagitis aguda, 132
espacio de muestreo, 70

espasmo esofágico, 132
especificidad, 31, 32, 86, 135-138, 143, 181, 186, 259
- de la búsqueda, 31
esperanza de vida, 358
esquizofrenia, 579
estadígrafos, 62
estadística, 108
- descriptiva, 22, 63, 65, 66, 246, 247, 456, 457
- inferencial, 24, 27, 64
- kappa, 108, 310
estándar de oro, 134, 180, 181, 182, 398
estenosis
- de la arteria renal, 129
- renal, 131
estimador, 71, 72, 86, 154, 212, 213, 245, 255, 262, 321, 457
- de precisión, 86, 123, 235, 236, 317, 398, 455, 457, 461, 469
estimación no sesgada, 212
estrategias
- de muestreo, 153
- observacionales, 245
estratificación, 71, 171, 173, 205, 210, 218, 228, 245, 260, 458
- pronóstica, 210
estratos, 56, 71, 73, 210, 230, 232, 256, 260, 446, 458
estructura
- de factores o dominios, 102
- de la propuesta de investigación, 149
- - anexos, 160
- - antecedentes, 150
- - - estado del arte, 150
- - - justificación, 151
- - - planteamiento del problema, 150
- - consideraciones éticas, 153
- - cronograma, 155
- - grupo de investigación, 154
- - - asistente de administración, 155
- - - asistente de investigación, 155
- - - coinvestigadores, 154
- - - expertos en estadística y otras metodologías específicas, 155
- - - investigador principal, 154
- - identificación, 150
- - metodología, 152
- - - diseño, 152
- - - instrumentos, 152
- - - muestra, 152
- - - plan de análisis, 153
- - - población, 152
- - - procedimientos, 153
- - - variables, 152
- - objetivos, 151
- - - específicos, 151
- - - general, 151
- - presupuesto, 157-159
- - referencias, 159
- - resultados esperados, 154
- - resumen, 150
- - título, 150
- de un estudio de cohortes, 34, 78, 179, 186, 227, 228, 247, 256
- PICO, 32
estudios, 34, 38, 59, 78, 85, 116, 118, 127, 170, 171, 173-176, 179, 182, 186, 194, 205, 227, 228, 229, 232, 245, 247, 248-252, 254-256, 258, 260, 261, 263-266, 301-303, 311, 338, 340, 372, 398, 462, 529, 542, 543
- analíticos, 188
- anidados de casos y controles, 232, 374, 375
- casos y controles, 176, 186, 250
- ciegos, 173, 205, 418, 566
- clínicos de un solo sitio, 411
- clínicos multicéntricos, 411
- de análisis económico, 170
- de cohorte, 34, 59, 78, 85, 115, 116, 118, 127, 178, 179, 186, 227, 228, 247-252, 254-256, 258, 260, 263-266, 462, 542
- de concordancia, 182, 186
- de conformidad, 182
- de consistencia, 182
- de correlación, 186
- de fase I, 368, 369
- de fase II, 369, 377, 411
- de fase III, 369, 377, 378
- de fase IV, 301, 369, 370, 377, 378, 398, 411
- de muestras, 61
- de prevalencia, 188
- de supervivencia, 123
- doble enmascarado, 173, 566
- enmascarado o ciego, 173, 566
- etnográficos, 184
- eventos adversos, 423
- - gravedad, 424
- - imputabilidad, 424
- - reporte, 424
- inclusión de pacientes, 421
- - aleatorización, 423
- - fuentes, 421
- - reclutamiento, 422
- integrativos, 38, 170, 182, 311, 542, 543
- longitudinal, 249, 250
- multicéntrico, 96, 171, 460, 581
- multisitio, 411

- negativo, 126
- observacionales, 171, 173-175, 194, 228, 245, 261, 264, 301-303, 338, 340, 372, 398, 529
- - analíticos, 228
- otros modelos, 182
- - análisis económico, 183
- - guías de práctica clínica, 183
- primarios, 542
- prospectivos, 266
- retrospectivos, 176, 229
- revisiones de la literatura y metaanálisis, 182
- secundarios, 182
- seguimiento, 423
- triples enmascarados, 173, 566
- ***v. t.*** **casos y controles, cohorte analítica,** correlación y concordancia, corte transversal, experimento clínico**,** sesgo
ética, 552
- comités de investigación y ética, 568
- comunidades menos desarrolladas, 563
- datos genéticos, 566
- experimentos clínicos, 564
- marco ético, 554
- - asociación colaborativa, 555
- - balance, 556
- - consentimiento informado, 556
- - independencia, 556
- - participantes, 556
- - respeto por los participantes/comunidades, 556
- - validez científica, 555
- - valor social, 555
- marco legal, 560
- - nacional, 560
- principios de la investigación, 553
- - beneficencia, 554
- - justicia, 554
- - respeto por las personas, 553
- sujetos especiales, 561
- ***v. t.*** **investigación**
evaluación, 25, 26, 62, 264, 456, 458
- ciega de los desenlaces, 546
- ciega o enmascarada del desenlace, 264
- crítica, 533
- de calidad de los artículos, 31, 36, 37, 307-310 540, 544, 550, 581, 582
- de escalas
- - apariencia, 101
- - confiabilidad, 109
- - - relacionada con el instrumento, 109
- - - relacionada con el tiempo de aplicación, 109
- - - relacionada con el sujeto, 110
- - contenido, 102
- - criterio, 106
- - - concurrente, 107, 108
- - - predictiva, 108
- - de constructo, 108
- - validez, 101
- de hipótesis 25, 26, 62, 264, 456, 458
- de tecnología, 386
- - conceptos básicos, 387
- - - actores, 388
- - - tecnología, 387
- - - tecnologías en salud, 387
- - definición, 388
- - dominios, 389
- - - aspectos éticos, 394
- - - aspectos legales, 395
- - - aspectos organizacionales, 395
- - - aspectos sociales, 394
- - - descripción y comparadores, 390
- - - efectividad clínica, 392
- - - impacto económico, 393
- - - problema y uso de la tecnología, 389
- - orientación y tipos de informes, 389
- - proceso, 395
- - - fase 1, 397
- - - fase 2, 397
- - - fase 3, 402
- - - fase 4, 404
- - - fase 5, 406
- - - fase 6, 406
- de tecnologías en salud, 386
- de una hipótesis o estimación de un IC, 86
- de validez externa, 550
- enmascarada de los desenlaces, 211
- económica, 349
- - costos en salud, 361
- - - clasificación, 362
- - - estimación de los costos, 361
- - - métodos, 364, 365
- - - secuencia de costeo, 364
- - definición, 351
- - objetivo, 349
- - oportunidad, 352
- - tipos, 352
- - - completas, 353
- - - de costo-beneficio, 360
- - - de costo-utilidad, 358
- - - parciales, 353
evento
- cardiovascular agudo, 264
- de tipo nominal, 275
- probabilísticos, 327
evidencia, 26, 34, 295, 305, 529-532, 536, 538, 541

- acumulada, 10, 303, 305
- asociación, 266
- búsqueda, 32, 35, 302, 307, 336, 397, 549
- - GPC, 34, 39, 332-334, 537
- - - consenso, 333
- - - recomendaciones, 334, 336, 342
- - literatura, 34-36, 305
- - metaanálisis, 34, 317
- calidad, 317-321, 331, 333, 338-341, 377, 380, 396, 400, 401, 532, 550
- certeza, 320
- científica, 35, 335, 592
- clara, 208, 377
- clínica, 258, 535, 538
- confusión, 458
- cuerpo, 320-322, 398, 400, 401
- de efectividad, 377
- disponible, 319, 339, 401
- documental, 363
- empírica, 91, 258, 358, 546
- estadísticas, 79
- evaluación, 26, 218, 300, 317, 338, 340, 405, 544
- fuentes, 35, 81, 183, 398, 454
- global, 317
- hipótesis, 25, 457
- indirecta, 319, 340, 401
- medicina (MBE), 4, 300, 326, 389, 529, 530, 533, 536, 537, 589
- mejor, 32, 34, 368, 376, 377, 400, 530, 534
- métodos, 301
- niveles, 340
- objetiva, 363
- proceso, 533
- - búsqueda de literatura, 534
- - evaluación de la información, 535
- - evaluación de resultados, 536, 537
- - práctica clínica, 536
- - pregunta, 534
- recogida, 370, 371, 379
- seguimiento, 253
- síntesis, 306, 319, 321, 339, 399, 400, 402-404
- sustancial, 34
- tablas, 339, 400
- tecnología, 386, 399
- toma de decisiones
- - elementos, 531
- - - consideraciones, 532
- - - experiencia individual, 531
- - - investigación, 532
- - - paciente, 532
- veracidad, 57
exactitud, 48, 58, 86, 96, 136-138, 213, 277, 412, 413, 429, 505, 548
examen
- crítico, 6
- diagnóstico, 23, 35, 390
experiencia clínica, 414, 529-531, 537
exclusión, criterios de, 67, 152, 536
experimentos
- aleatorizados (ECA), 123, 168, 171, 200, 211, 218, 301, 373, 542, 563
- - doble enmascarado, 171
- análisis, 216
- - comparación de base, 216
- - comparación de resultados, 217
- - - ajuste, 218
- - secuencial, 202
- asignación
- - de los sujetos, 208
- - - aleatorización, 208
- - - estratificación, 210
- - enmascarada, 211
- - - diseño, 204
- - - evaluación, 214
- - por bloques, 222
- - por grupos, 202
- - simple, 221
- aspectos éticos, 220
- científico, 162
- controlado, 68, 204, 211, 374
- convencional, 200
- cruzado, 200, 208, 375
- de asignación por grupos, 202, 208
- de descontinuación, 203
- de equivalencia, 201, 202, 208
- de intención de tratar, 217
- de manejo, 217
- de no inferioridad, 202
- explicativo, 217
- factorial, 202
- ficticio, 213
- individual (N = 1), 204
- Magpie, 117-121, 124-127
- megaexperimentos, 203
- natural, 198
- pragmático, 217, 377, 378
- simple, 573
- clínico, 8, 34, 68, 76, 78, 123, 168, 170, 171, 173, 186, 198-202, 204, 209, 211, 213, 218, 220, 221, 301, 373-375, 377, 378, 414, 416, 424, 426, 427, 429, 434, 459, 542, 548, 563-566, 573, 577
- conducción, 244
- intervenciones, 213
- pregunta, 204

- - criterios
- - - de exclusión, 206
- - - de inclusión, 205
- - población, 205
- - tamaño de la muestra, 208
- racionalidad del ECA, 211
- pragmático, o de intención de tratar, 217
exposición
- a fármacos, 242
- al factor de interés, 117
- de control, 23, 67, 199, 244, 248
- efecto de la, 122, 259, 261, 372
- estatus de, 375
- experimental, 117, 245
- factor de, 23, 68, 84, 116, 117, 122, 177, 179, 226, 227, 229, 236, 239, 244, 245, 248, 253, 256, 258, 264, 265, 350
- frecuencia de, 68, 84, 85, 121, 199, 372
- grado de, 257, 259
- grupo de, 69, 259, 260, 264, 265
- intrauterina, 227
- variable de, 53, 55, 84, 86-88, 232, 245, 265
- ventana de, 374, 375
- *v. t.* **componentes de un experimento, tipos de** experimentos
extracción de datos, 399, 561
extrapolación, 64, 66, 67, 205, 213, 219, 255, 257

F

factibilidad, 25, 61, 77, 147, 148, 203, 292, 334, 374, 402, 415, 431
factor
- de confusión, 55, 57, 236, 240, 248, 260, 458, 536
- de corrección del sesgo, 277
- de expansión, 156
- de exposición, 23, 248, 249, 258
- de buen pronóstico, 246
- de mal pronóstico, 246
- de riesgo, 23, 32, 121, 169, 176-179, 190, 236-238, 240, 245, 247, 248, 250, 256, 257, 264, 337, 339
- externo, 245
- falso, 80, 131, 132, 134-139, 143, 228, 239, 258, 259, 374, 541, 543, 579
- negativo, 131, 132, 134-136, 138, 139, 228, 374
- positivo, 80, 131, 135-139, 143, 228, 239, 258, 259
- pronósticos, 246
- protector, 246
falacia
- de la inducción, 16
- de la inducción, o afirmación del consecuente, 66
- formales, 17
- informales, 17
falla
- cardíaca, 381
- terapéutica, 218
falsabilidad, 26, 65
falsada, 25
falso
- negativo, 126, 133, 134
- positivo, 133
falta
- absoluta de recursos, 435
- de redes de apoyo, 436
farmacoeconomía (FC), 367, 368, 370, 380
- - análisis, 380
- - análisis económico, 376
- - principios y tipos de estudios, 378
farmacoepidemiología (FE), 370
- diseños de investigación, 373
- diseños para evaluar el riesgo, 373
- sesgos, 371
- - de confusión, 372
- - de información (mala clasificación), 372
- - de selección, 371
fármaco-genómica, 567
farmacovigilancia 38, 301, 370
fenotipo, 11
fibromialgia, 170
fibrosis pulmonar, 456
flexible, 14
formato, 8, 99, 101, 153, 159, 160, 165, 283-285, 333, 405, 406, 414, 416, 417, 427, 429, 452, 468, 469, 485-487, 492, 498-501, 503, 582
- de recolección, 8
formulación de preguntas, 11, 18, 21, 23, 447, 448
formularios de captura, 166, 216, 312, 425, 570
fracción de riesgo atribuible, 265
frecuencia, 139
- porcentual, 506
fricción de costos, 363
fronteras, 515
- de las clases, 515
fuentes de error, 257
- sesgos en la selección, 257
- - desbalance, 259
- - en el muestreo, 257
- - en la asignación, 258

- - mala clasificación, 258
función de riesgo, 372

G

generalizabilidad, 154, 203, 206, 207, **221,** 242, 283, 289, 535, 536, 544, 549, 569
generalización, 6, 13, 66, 69, 75, 205, 207, 214, 234, 291, 295, 456
genotipo, 11
grado
- de seguridad débil, 341
- de seguridad fuerte, 341
- de severidad moderado, 50
gráficos, 163, 506, 508-511, 513-517, 519, 520, 522-527
- de barras agrupadas, 522
- de barras simples, 508, 520
- de Bland y Altman, 277
- de bosque, 315
- de caja y bigote (***Box plot***), 512, 517
- de dispersión, 523, 524
- de embudo, 315, 316, 319
- de estrella, 527
- de línea, 519, 520, 523
- de radar, 525
- de sectores, 509
- de telaraña, 525
- y métodos multivariados, 236, 245, 265, 458
- ***v. t.*** **diagramas**
gastritis, 370
Guías de Buena Práctica Clínica, 560
glomerulonefritis postestreptocóccica (GPE), 65, 69
glomerulopatía diabética, 57, 261
glucemia en el diabético, 132
grupo
- control, 83, 117, 171-173, 175, 176, 200, 201, 208, 209, 211, 213, 214, 216, 220, 238, 239, 256, 321, 372, 417, 531, 548, 550, 563, 564, 566
- experimental, 201, 208, 209, 214, 216, 542
- focal, 185, 285-287, 289
guías de práctica clínica, 39, 183, 185, 300, 331, 333, 338, 339, 364, 379, 400, 412, 534, 538, 542, 560
- definición, ventajas y desventajas, 331
- desarrollo
- - métodos, 332, 379
- - - consenso
- - - - basadas en evidencia, 39, 333
- - - - formal, 333
- - - - informal, 332
- herramientas, 342
- pasos, 334
- - búsqueda de la evidencia, 336
- - constitución del grupo, 335
- - - conflictos de interés, 336
- - evaluación de la evidencia, 338
- - implementación, 341
- - información, 339
- - recomendaciones, 339
- - selección del foco, 334
- ***v. t.*** **diseño**

H

hemofilia, 234
hemorragia cerebral, 215
hepatitis
- B (HBV), 137
- C, 368
hernia, 424
heterogeneidad, 71, 72, 302, 303, 305, 311-315, 317-319, 339, 340, 400
- evaluación, 305, 312
hipercolesterolemia, 79, 143, 207, 264
hiperglucemia, 62
hipertensión, 34, 131, 182, 215, 264, 378, 447
- arterial (HTA), 27, 32, 33, 57, 131, 182, 183, 206, 246, 380, 460, 557
- crónica, 530
- estadio 1, 49, 173
- estadio 2, 49, 173
- sistólica, 182
- - aislada, 33
hipertrigliceridemia severa, 207
hipofisectomía en retinopatía diabética, 544
hipoglucemia, 264
hipótesis, 5, 8, 9, 11, 15-17, 24-29, 64-67, 74, 78, 79, 108, 111, 125, 164, 174, 188, 195, 204, 205, 216, 219, 239, 260, 261, 271, 283, 294, 319, 371, 452, 534, 548, 569, 570
- a dos colas, 83, 202
- a una cola, 83, 202, 271
- alterna, 27, 80, 81, 457
- características, 25
- científica, 5, 24, 27
- de asociación, 244
- de interés, 83
- nula, 457
- - de no diferencia, 239
- evaluación, 26, 53, 62, 66, 67, 75, 82, 86, 204, 255, 264, 454, 456-458
- fuentes de, 174, 175, 454
- generación de, 28, 75, 189, 193, 194, 282, 294

- nula, 80, 84, 85, 104, 235, 239, 318, 457
- temprana, 28
- veracidad de la, 25, 26, 283
histograma, 517, 521, 527
historia clínica, 129
hoja de cálculo, 155, 159, 467
homogeneidad, 71, 72, 76, 79, 109, 207, 287, 312, 317, 400, 513, 525
hospitalización, 420

I

identificación, 469
impétigo, 66, 68
importancia relativa, 394
incertidumbre, 6, 25, 28, 55, 60, 76, 77, 129, 132, 133, 213, 264, 320, 325, 328-331, 334, 350, 356, 359, 360, 377, 380, 381, 391, 403, 531, 544, 564-566
- nivel de, 360
- total, 329
- umbral de, 132, 133
incidencia, 19, 26, 34, 65, 68, 69, 114-119, 122-124, 127, 174, 175, 180, 190, 192, 215, 231, 241, 247-249, 251-253, 256, 258-263, 265, 266, 270, 279, 363, 373, 375, 391, 456, 462, 544, 578, 582
- acumulada, 247, 252
- acumulativa, 115
infarto 143, 238, 281, 338, 509
- agudo de miocardio (IAM), 132, 169, 172, 381
- cardíaco, 169, 170, 183, 424, 529
- cerebral, 183
- de miocardio, 85, 203, 246, 262, 382, 388, 458, 532, 544, 545
- miocárdicos, 215
infección
- de piel (impétigo), 65
- nosocomial, 116
- por VIH, 436
- respiratoria aguda, 262
- urinaria, 18
- - no complicada, 547
inferencia, 5, 6, 15-17, 19, 20, 22-25, 27, 28, 57, 62, 64, 66, 69, 75, 76, 189, 194, 226, 257, 370, 371, 373, 398, 457, 458, 564
influencia viral, 34
informática biomédica, 588
- fundamentos, 588
- interacción, 589
- - epidemiología, informática biomédica y docencia, 592
- - informática y epidemiología, 589
- proyección, 593
- uso, 592
incapacidad persistente o significativa, 420
incertidumbre, 328, 566
inclusión, 70
- criterios de, 152, 536
independencia, 237
índice
- de correlación y concordancia, 279
- de repetibilidad, 276
- de sesgo, 271
inducción analítica, 292
inferencia, 19, 62
- deductiva, 16
- inductiva, 16, 66
influenza viral, 34
información, 19, 371, 533
informes primarios de investigación, 36
instrumento
- de medición, 58, 59, 95, 97, 100, 563
- de recolección, 417, 419
interacción, 11, 12, 24, 56, 92, 152, 184, 195, 196, 228, 232, 242, 244, 248, 266, 283, 285-288, 295, 428, 432, 437, 445, 455, 458, 459, 462, 532, 565, 572, 589-593
insuficiencia
- cardíaca, 183, 530
- - congestiva (ICC), 137, 434
- - severa, 206
- renal crónica terminal, 27
intercambio temporal, 359
interevaluador, 109, 110
internet, 33, 35, 36, 39, 41, 98, 126, 222, 307, 338, 531, 534, 585, 589
interrogatorio, 129
intervalo
- de clase, 515
- de confianza, 60, 124, 263, 321, 340, 548
- muestral, 70
- y razón, 52
intervención, 4, 7-9, 19, 23, 32, 34, 35, 61, 65, 68, 73, 76, 79-81, 91, 92, 109-111, 113, 116-119, 121, 126, 127, 131, 151-153, 158, 163, 165, 166, 168-175, 183, 186, 198-209, 211-222, 245, 246, 259, 268, 300-314, 316, 317, 319-322, 327, 331, 334, 337, 340, 349-357, 360-365, 376-383, 387-390, 397-403, 405, 406, 414, 418, 423, 424, 428, 434-438, 451, 529-537, 541, 545-550, 552, 557-559, 561, 563-567, 569, 573, 591
intuición, 5, 529, 530, 538
investigación, 3
- científica, 19, 24, 61, 562, 565

- conocimiento
- - adquisición, 4
- - autoridad, 4
- - ensayo y error, 5
- - intuición, 5
- - razonamiento deductivo, 5
- - razonamiento inductivo, 5
- - razonamiento lógico, 5
- - tradición, 4
- cualitativa, 20, 73-75, 94, 281, 282, 284, 288-290, 292-295, 394, 399
- cuantitativa, 20, 55, 62, 75, 282, 288
- de control, 534
- en salud, 3
- - definición, 3
- - tipos, 7
- - - básica, 7
- - - clínica, 7
- - - epidemiológica, 7
- - - servicios de salud, 7
- método científico, 6
- - fases y pasos, 7
- - - análisis, 8
- - - comunicación, 9
- - - concepción, 8
- - - implementación, 8
- - - planificación, 8
- primaria 170, 171, 398, 403, 404
- secundaria, 170
- *v. t.* **diseño, ética, pregunta, propuesta de** investigación, protocolo
isquemia, 137

J

juicio, 31, 57, 183, 219, 289, 293, 310, 320, 452, 530, 535, 544, 549, 554, 574, 577
justicia, 189, 221, 394, 553, 554, 573, 576

K

kappa ponderado, 272-275

L

labio leporino, 241
laringotraqueitis, 314, 316
lectura crítica, 166, 534, 541, 550, 551
lenguaje prescriptivo, 15
lepra, 529
leucemia, 173, 253
- linfoide aguda (LLA), 251-253
- mieloide aguda, 362
limitación de la validez externa, 233
límites de concordancia, 276
literatura
- científica, 36
- - búsqueda, 36
- - - artículos, 40
- - - bases de datos latinoamericanas, 40
- - - EMBASE, 38
- - - literatura gris, 41
- - - otras herramientas, 38
- - - PubMed, 37
- médica, 30
- - búsqueda, 30
- - - objetivos, 32
- - - pregunta, 34
- - - proceso, 32
- - fuente de información, 35
- - - colegas, 35
- - - fuentes bibliográficas, 36
- - - personal, 35
locus, 108
lógicas binarias, 48

M

malaria, 4
magnitudes, 457
mala clasificación
- diferencial, 258
- no diferencial, 258
manejo
- centralizado, 420, 421
- de datos, 222, 340, 417, 419, 431
- no centralizado, 420
manual de códigos, 460
marca de clase, 517
marco muestral, 70-72
mediana, 515
medicina basada
- en apariencias, 533
- en conferencias, 533
- en efervescencia, 533
- en emergencias, 533
- en eminencia, 533
- en evidencia, 326, 529-531, 533, 535, 537
- en experiencia, 533
- en gerencia, 533
- en grandilocuencia, 533
- en jurisprudencia, 533
- en prepotencia, 533
- en vehemencia, 533
- *v. t.* **evidencia**

medición, 48
- atributos, 57
- cualitativa, 48, 399
- cuantitativa, 48, 399
- de los costos, 329, 358, 361, 378, 391, 393, 543
- fuente de error, 58
- tipo de error, 59
- - sesgo, 59
- variabilidad, 57
- ***v. t.*** **enfoque de la medición**
medidas, 113, 190, 241, 314, 373, 462, 513
- de asociación, 113, 116, 122, 190, 241, 314, 373, 462, 458
- - número necesario para tratar, 121
- - reducción del riesgo, 118
- - riesgo relativo (RR), 119
- - riesgo relativo indirecto, 119
- de dispersión, 513
- de efectividad, 172, 317, 353, 355, 358, 377-379, 393, 397-399, 401-403, 405
- de frecuencia, 113, 114
- - incidencia, 115
- - prevalencia, 114
- - usos, 116
- de impacto, 113
- - razones y proporciones, 113
- de precisión, 123, 461
- - calculadoras, 126
- razón de "peligros" (***hazard ratio***)**, 123**
- razón de prevalencias, 122
medio electrónico, 38, 40, 101, 150, 307, 406, 427, 436, 590
membresía graduada, 50
meningitis bacteriana, 214, 263
mensurables, 92
méritos, 465, 476, 574, 580, 584
metaanálisis, 20, 32, 34, 36-38, 137, 182, 186, 299-304, 311-315, 317, 319, 321, 326, 339, 358, 399, 400, 436, 437, 532, 535, 538, 542
- de los datos, 312
método
- científico, 3, 6, 19
- de capital humano, 363
- cualitativos, 183
- - documentos, análisis de textos o discursos, 185
- - entrevistas, 185
- - - grupales (grupos focales), 185
- - - individuales, 185
- - observación, 184
- para evaluación de concordancia, 269
- - eventos de tipo ordinal, 272
- - kappa ajustado, 272
- - variables categóricas, 269
- - - eventos nominales, 269
migraña, 201, 424
minimización de costos, 378
miopía, 174
modalidades según el sujeto de estudio, ***v.*** **sujeto de estudio**
modelado, 464, 465
- captura de los datos, 473
- - componentes, 465, 474
- - - atributos, 467
- - - entidades, 466
- - - patrones útiles, 476-479
- - - relaciones, 470
- - patrones útiles, 470
- - - relaciones, 471
- - - valores de referencia, 472
modelo
- de captura, 481
- de datos, 481
- de efectos aleatorios, 313
- de efectos fijos, 313
- regresión, 86-88
- - lineal múltiple con variable de exposición continua, 87
- - lineal simple con variable de exposición continua, 86
- - lineal simple con variable de exposición dicotómica, 86
- - logística múltiple, 88
- - logística simple con variable de exposición continua, 88
- - logística simple con variable de exposición dicotómico, 88
modificadores de efecto, 459
*modus tollens***, 28**
momios, suertes u ***odds***, 119
monismo, 11
monitor, 413
- funciones, 417, 423, 425-432, 460
monitorización, 425
- basada en riesgo, 432
- estrategias, 427
- información, 425
- objetivos, 425
- visitas, 428, 429
muestra, 62, 67, 76, 78, 212
- aleatoria, 68, 69, 78, 123, 213, 314, 374
- - simple, 71, 72
- atípica, 60
- error tipo I y nivel de confianza, 80
- error tipo II y poder del estudio, 81

- índice de adecuación, 103
- inferencia, 64, 66
- no aleatorias, 74
- no sesgada, 60, 237
- obtenida, 59
- parámetros, 78
- población, 59, 61, 62, 70, 76, 114, 189, 193, 206, 265, 314, 454, 456, 458, 507
- - blanco, 61, 63, 75
- polietápica, 80
- precisión, 63
- preguntas, 74
- presentación gráfica, 314
- probabilística o muestra de conveniencia, 463
- representativa, 64, 66, 67, 73, 110, 245, 542
- resultado, 64
- selección, 9, 68, 70, 75, 77, 114, 156, 420, 461
- válida, 63, 74, 108
- valores, 78
- - estadísticos, 78
- variabilidad, 78
- ***v. t.*** **muestreo, obtención de muestra, tamaño** de muestra
muestreo, 67, 69, 74
- aleatorio simple, 70-72, 80
- bietápico, 80
- de conveniencia, 69
- de no afijación proporcional, 73- en estudios cualitativos, 289-291
- error, 78
- estratificado, 71, 72
- no probabilístico, 75
- por densidad de incidencia, 115, 231
- por grupo de riesgo, 231
- probabilístico, 69
- - selección, 70
- - aleatorio simple, 70-72, 80
- - sistemático, 70
- - estratificado, 71
- - por conglomerados, 71, 72
- - - bola de nieve, 74
- - - no probabilístico, 72, 73, 256
- - - polietápico, 72
- - - por conveniencia, 73, 291
- - - por cuotas, 73
- - - por juicios, 73
- - - probabilístico, 69-73
- - - sistemático, 70, 71
- proceso, 74
- ***v. t.*** **muestra, obtención de muestra, tamaño** de muestra
mundos o planos de realidad, 45, 46

N

nefritis intersticial por antiinflamatorios no esteroideos (AINE), 178
nefropatía diabética, 27, 261
negociación de tiempo (time trade-off), 359
neoplasia maligna, 56
neumonía, 16, 34, 65, 70, 172, 199, 200, 219, 275
- bacteriana, 16, 172
- nosocomial, 115
nivel
- de confianza, 80
- de expuesto, 375
- de no expuesto, 375
- de medición para variables, 49
- de significación 240
nodo
- de decisión, 327
- terminal, 327
nomos, 25
número
- aleatorio, 70, 222
- necesario para tratar (NNT), 118, 121, 265, 400, 550

O

obesidad, 88, 264
objetivos, 14
- primarios, 151
- secundarios, 151
oblimin, 105
observación
- estructurada, 283
- no estructurada, 283
- no participante, 185, 282
- participante, 185, 282, 283
obtención de muestra, 72
- conglomerados, 72
- no probabilístico, 73
- - bola de nieve, 74
- - por conveniencia, 73
- - por cuotas, 73
- - por juicios, 73
- polietápica, 72, 80
octiles, 515
odds ratio, **84, 118, 120, 176, 192, 226, 227,** 400, 462
oferta, 349
operacionalización en contexto, 93
operaciones, 22
orden variable según la conveniencia, 577

ordinal, 275
organización, 388
otitis media, 24, 25
- aguda, 23, 24
outcome, 32

P

pancreatitis, 269, 275
- hemorrágica, 174
papel analítico, 47
paradigma, 294
paradoja de Simpson, 228
parámetro poblacional, 62, 74, 212
- común, 457
paro cardiorrespiratorio, 264
parotiditis, 541
partes de una tabla, 505
participantes, 467, 483
patrón
- de diseño, 470
- de oro, 57, 96, 107, 108, 133-136, 143, 166, 268, 269, 273, 542
pérdida
- al seguimiento, 81
- diferencial, 546
pericarditis tuberculosa, 202
periodontitis, 557
persistencia, 435
perspectiva del análisis, 403
pesos lineales, 273
pictogramas, 508, 512
pielonefritis, 213
- aguda, 212
pirámides poblacionales, 521, 527
pitiriasis rosada, 201
placebo, 23, 68, 79, 117-119, 121, 124, 126, 172, 200, 202, 203, 214, 220, 315, 321, 570
- comparación, 337, 339, 404, 434
- efecto, 215, 216, 417, 544, 546, 555, 564-566
plan de análisis, 454
- ***a priori,*** 456
- - hipótesis, 456
- - - analíticas, 457
- - - descriptivas, 456
- - objetivos, 456
- - población, 458
- - tipo, 457
- - variables, 459
- análisis final, 461
- - exploración, 462
- - - pasos, 462, 463
- - mediciones crudas, 461
- análisis inicial, 461
- información, 459
- - procesamiento, 459-461
- - recolección de datos, 460
población, 62, 534
- abstracta, o de referencia, 61
- accesible, 64, 244
- blanco, 61-67, 74, 254, 457
- - elegible, 67
- - específica, 456
- - de estudio, 23, 63, 64, 66, 76, 77, 79, 86, 98, 111, 205, 255, 257, 305, 415, 420, 431, 456, 458, 461, 534, 556, 573
- - de referencia 51, 65-67, 74, 75, 230, 235, 236, 255, 456
- - muestra, 62, 64
- - parámetros, 63
- - - control (intervención de control), 65
- - - desenlace, 65
- - - dominio, 65
- - - exposición (intervención), 65
- - - población, 65
- - precisión, 63
- - prueba, e inferencia, 65
- - referencia, 65
- - selección y muestreo, 67, 68
- - validez, 63
- - vulnerable, 221
- control, 32
- de estudio, 23, 255
- de interés, 61, 399
- de referencia, 65, 255, 456
- fuente, 255
- objetivo o diana, 62
poder, 26, 28, 69, 77, 81, 83-88, 94, 164, 170, 190, 202, 205, 208, 236, 239, 240, 283, 285, 286, 313, 374, 417, 437, 458, 478, 541, 548, 549, 573, 592
- del estudio, 69, 81, 164, 190, 236, 239, 458
- estadístico bajo, 374
polidipsia, 139
polígono de frecuencias, 517, 519
polítomas, 48
poliuria, 139
porcentajes, 519
poscodificación, 460
potencia estadística, 79
pragmáticas, 18
precodificación, 460
predicciones, 19, 20, 64, 96
preeclampsia, 117, 121, 124, 204
preferencia, 91, 184, 326, 328, 332, 333, 336, 341, 350, 354, 359, 360, 379-381, 394, 398,

403, 404, 530, 543, 565
- de tiempo, 18, 210, 271, 310, 354, 355, 358-360, 379
pregunta(s), 11, 17, 62
- abierta, 291, 445, 449
- adaptación, 95-97, 99, 387
- atributos, 21
- - amplia, 21
- - contestable, 21
- - específica, 21
- - éticamente apropiada, 21
- - importante, 21
- - interesante, 21
- características, 25
- - compatibilidad, 26
- - evaluabilidad, 26
- - poder predictivo, 26
- - relevancia, 26
- - simplicidad, 26
- cerrada, 445, 459, 450
- clases, 19
- - descriptivas, 20
- - inferenciales, 20
- - - explicativo, 20
- - - pragmático, 20
- clínica, 31, 32, 35, 37, 186, 543
- cualitativas, 20, 71
- cuantitativas, 20
- de investigación, 11, 17, 21, 23, 25, 27, 30, 34-36, 53, 57, 62, 65, 77, 78, 83, 94, 148-151, 161, 163, 164, 167-169, 189, 190, 201, 204, 256, 266, 283, 287, 288, 301, 304, 305, 312, 398, 404, 412, 454, 456, 457, 463, 464, 466, 467, 502, 526, 534, 541, 542, 545
- enfoque en el problema, 18
- enfoque en el tópico, 18
- descriptivas, 62, 456
- escritura de, 15, 161, 209, 309, 317, 406, 446, 447, 449, 450, 582
- evaluación, 26
- formulación, 21
- - estrategias, 21-23
- hipótesis, 24
- ***v. t.*** **conocimiento científico, construcción de** preguntas, investigación
prehipertensión, 49
premisas, 5, 15
preparación de datos, 492
- estructura, 492
- exportación de los datos, 499
- transformación, 492
- vinculación
- - campos del tipo fecha, 498
- - datos de otros registros, 498
- - otras tablas, 496
- - valores, 496
presentación estadística de datos, 459, 502
- *v. t.* **datos estadísticos**
prevalencia, 19, 62, 77, 82, 84, 86, 88, 114-116, 122-124, 131, 139, 141, 142, 144, 172, 175, 188-190, 192, 194, 227, 240, 248, 271, 272, 275, 456, 457, 461, 581
- anual, 115
- de interés en la población de estudio, 86
- de periodo, 115
- de punto, 115
- estudios de, 77, 114, 116, 188, 189
primum non nocere, **554**
principio(s)
- comparabilidad, 234, 379
- de eficiencia, 236
- de equilibrio clínico, 220
- de la base del estudio, 235
- de precisión comparable, 235,236
- incertidumbre clínica, 564
- incertidumbre terapéutica, 564
- indeterminación clínica, 564
- selección de controles, 234
probabilidad
- preexamen, 132, 139-144
- de ocurrencia, 116, 328, 329, 362, 363, 380
- intermedia, 142
problemas de la autoría, 573
- causas, 580
- orígenes, 573
- - aumento de la actividad, 573
- - autocitación, 579
- - difusión de criterios, 573
- - exclusión injustificada, 576
- - inclusión injustificada, 574-576
- - omisión de información, 577
- - orden inapropiado de los autores, 576
- - publicaciones repetidas, 578
- - uso incorrecto de los criterios, 574
- ***v. t.*** **autoría**
proceso
- de adaptación, 95-97
- de desarrollo y aprobación del uso de medicamentos, 368
procedimiento para analizar datos, 288
promax, 105
pronóstico, 7, 19, 23, 32, 34, 132, 169, 172, 178, 209, 211, 245, 246, 248, 261-263, 265, 266, 301, 305, 322, 337, 350, 372, 423, 434, 531, 541, 542, 545-547, 566, 591
propiedad intelectual, 573, 585

proporción, 9, 30, 40, 52, 55, 62, 63, 66, 73, 81-84, 86, 88, 106, 113-115, 118-120, 125, 170, 190, 200, 205, 213, 215, 239, 247, 252, 253, 262, 263, 268, 270, 271, 276, 308, 318, 319, 381, 506, 507, 559, 581, 586
- de incidencia, 115, 252
propuesta de investigación, 9, 25, 49, 65, 147-154, 157, 159-161, 163, 166, 208, 221, 250, 276, 294, 301, 303, 317, 400, 419, 530, 533, 538, 548, 568, 583, 584
- antepropuesta, 148
- herramientas para elaborar un presupuesto, 159
- propuesta exitosa, 160
- ***v. t.*** **estructura de la propuesta de investigación,** investigación
protocolo, 162
- estructura, 163
- presentación, 163
- - resumen, 163
- - texto, 163
- - - aspectos éticos, 165
- - - cronograma, 165
- - - documentos anexos, 165
- - - introducción, 163
- - - justificación, 164
- - - marco teórico, 163
- - - métodos, 164
- - - objetivos e hipótesis, 164
- - - planteamiento del problema, 164
- - - presupuesto, 165
- - título del proyecto, 163
- problemas comunes, 166
- recomendaciones, 166
- ***v. t.*** **investigación**
proyecto de investigación, 162
prueba(s)
- de hipótesis, 271, 457
- diagnóstica, 129
- - características operativas, 132
- - curvas (COR), 137
- - de concordancia, 110, 182
- - de Friedman o coeficiente, 110
- - de significación estadística, 125
- - diagnóstica, 32, 86, 130-133, 170, 186, 268, 300, 302, 305, 306, 338, 402, 550
- - en paralelo, 19, 143
- - en serie, 143
- - especificidad, 135
- - - indicaciones, 136
- - - relación, 136
- - evaluación, 133
- - ideal, 130
- - interpretación, 138
- - múltiples, 143
- - patrón de oro, 133
- - potencia de la, 81
- - razones de probabilidad, 141
- - - interpretación, 142
- - - positivas y negativas, 142
- - - uso, 142
- - sensibilidad, 134
- - ***t*** pareada, 110, 111
- - usos, 131
- - valores predictivos, 139
- - - resultado negativo, 140
- - - resultado positivo, 140
pseudoexperimento, 199
psoriasis, 201
publicaciones vivas, 302, 303
punto de corte, 133, 137, 138, 141, 144, 199, 306

Q

QALY, 359, 360, 379, 381-383, 403
quartimax, 105

R

raciocinio lógico, 14
racionalidad, 326
rango, 51
- intercuartílico, 513
razonamiento
- deductivo, 5
- inductivo, 5, 6, 16
- lógico, 5
razón, 182
- de controles a casos, 239
- de costo-efectividad, 355-357, 379, 382
- de prevalencias, 122
- de probabilidades, 86, 141-143, 227
- - negativas, 142
- - positivas, 142
- de riesgo, 190
- - relativo, 190
- de suertes, 118, 120, 121, 192, 193, 227, 241
- incremental de costo efectividad, 354, 398, 403
recolección
- de datos, 65, 78, 178, 229, 287, 445, 456, 543, 590
- base de datos, 417
- formato, 416, 455
- tablas, 484, 495
- de información, 416
- - aleatorización, 418
- - control, 420

- - enmascaramiento, 418
- - estandarización, 418
- - - entrenamiento de participantes, 418
- - estudio piloto, 419
- - formatos, 416
- - manuales de procedimientos, 419
- - proceso y análisis de datos, 420
red nomológica, 93, 94
reducción
- absoluta del riesgo (RAR), 118, 219, 400
- relativa de riesgo, 119, 532
referencias
- no ubicadas, 309
- obtenidas, 309
- solicitadas, 309
reflujo gastroesofágico, 132
refutabilidad, 25
refutar hipótesis, 371
registro
- de control de medicamentos, 420
- de tamizado, 420
regla de validación, 486, 487, 489, 490
regresión al promedio, 199, 200
relaciones
- de novo, 5
- entre variables, 53
- - desenlace, 55
- - exposición, 53
- - - asociación, 55, 56
repetibilidad, 130
repetitividad, 544
representación, 327
- balance del árbol, 327
- nodo de decisión, 327
- nodos probabilísticos, 328
- simetría del árbol, 327
representatividad de la muestra, 189, 212
restricción de información, 559
resultados de investigación primaria, 398
retinopatía diabética, 544
retraso mental, 206
revisión sistemática y metaanálisis, 33, 37, 163, 166, 182, 299, 302, 303, 305, 307, 314, 315, 317, 319, 326, 333, 354, 438
- aportes y limitaciones, 302
- búsqueda, 307
- - estudios no publicados, 308
- - estudios publicados, 307
- - manejo, 309
- calidad de evidencia, 317, 318
- - tablas, 321
- definición, 300
- narrativa, 301
- pasos, 304
- - pregunta de investigación, 304
- - - condición que se va a estudiar, 306
- - - criterios de inclusión y de exclusión, 305
- - - desenlaces de interés, 306
- - - diseño y calidad metodológica, 307
- - - intervención que se va a estudiar, 306
- - - población, 305,
- selección de los estudios, 310
- - análisis de la información, 312
- - calidad, 311
- - heterogeneidad, 312
- - métodos de extracción de los datos, 312
- - presentación gráfica, 314
- - sensibilidad y de subgrupos, 316
- - sesgo de publicación, 315
- - síntesis de resultados, 313
- sistemática de la literatura, 33, 182, 326
- tipos de revisiones sistemáticas de literatura (RSL), 300
riesgo
- absoluto, 179, 196
- atribuible, 265
- basal, 550
- de desarrollo del evento, 375
- del evento, 257
- mínimo, 153, 557, 559
- relativo, 55, 56, 68, 84, 85, 118-120, 126, 176, 177, 196, 219, 226, 241, 242, 256, 262, 263, 265, 314-316, 321, 373, 400, 462, 544, 548
- - indirecto (***odds ratio***), 84, 462
- - reducción, 119, 532
- superior al mínimo, 557
ritmo de galope, 137
RR indirecto, 227
rubeola, 227, 541

S

saltos condicionales, 474, 475, 482
sangrado gastrointestinal, 372
sarampión, 541
sedentarismo, 264
seguimiento, 68, 81, 115, 116, 119, 122, 123, 131, 132, 152, 153, 155, 156, 159, 165-167, 179, 193, 195, 200, 203, 205, 208, 215, 216, 218, 227, 231, 232, 247, 249-254, 255, 256, 260, 263, 266, 310, 311, 318, 321, 326, 335, 339, 340, 371, 374, 376, 380-383, 388, 398, 404, 412-416, 419, 421-423, 426, 429, 430, 432, 435, 451, 455, 460, 542, 547, 556, 563, 570
- de casos, 132

- prolectivo, 250
- prospectivo, 229, 249, 250
- retrolectivo, 250
- retrospectivo, 176, 229, 250
selección, 67, 371
sensibilidad, 30-33, 35, 37, 48, 81, 86, 96-99, 110, 111, 129, 134-144, 170, 180, 181, 186, 195, 196, 219, 258, 259, 263, 268, 285, 302, 305, 312, 316, 377, 380, 381, 383, 403, 455
- de la búsqueda, 30
sesgo, 45, 47, 49, 51, 53, 55, 57-59, 62, 100, 130, 195, 196, 211, 213-216, 239, 253, 257, 259, 263, 272, 277, 307, 311, 317, 318, 372, 535, 542-547
- ausencia de, 255
- de asignación, 245, 258, 259
- de aversión a los extremos, 100
- de cointervención, 214
- de confusión, 245, 372
- de detección temprana (***lead-time bias***), 253, 254
- de información, 236
- de mala clasificación, 233, 235, 238, 239, 258, 259, 263
- - diferencias, 258
- de medición, 238, 239, 264
- de muestreo, 258
- de Neyman, 228
- de publicación, 304, 305, 308, 315, 317, 319, 340, 401
- de recuerdo, 227, 231-233, 237, 238, 530
- de selección, 230, 232, 234-239, 257, 371
- de sobrevida (o de Neyman), 231
- de supervivencia, 234
- del adelanto del tiempo del diagnóstico, 253
- del recuerdo, 231
- efecto del, 271
- factor de corrección del, 277
- hacia la no diferencia, 259
- índice del, 271, 272
- por mala clasificación no diferencial, 259
- protopático, 373
- riesgo de, 340, 399, 401
significación
- clínica, 208, 283, 535, 536
- estadística, 69, 81, 125, 126, 208, 535, 536
síndrome
- coronario agudo, 207
- de choque séptico, 238
- de dificultad respiratoria aguda, 170
- de Down, 568
- depresivo, 102
- respiratorio agudo severo, 4
sinovitis
- clínica, 206
- inflamatoria persistente, 206
síntesis
- cuantitativa, 312, 314, 400
sistema de doble digitación, 461
sobreapareamiento, 257
sordera, 263
- neurosensorial posmeningitis, 263
subgrupos, 462
subjetivo, 11, 13
sujeto de estudio
- concordancia, 180
- - características operativas, 180
- - correlación, 180
- - tendencia, de correlación y de concordancia, 181
- diseños analíticos, 175
- - casos y controles, 175
- - - cohortes, 178
- diseños descriptivos, 174
- - reporte de un caso, 174
- - serie de casos, 174
- estudios observacionales, 173
- experimental, 170
- - enmascaramiento o cegamiento, 173
- - estratificación, 173
- - grupo control, 171
- - intervención de control, 172
- - limitaciones, 173
- - placebo, 172
- - sujetos, 172
- investigación primaria, 170
- otros estudios, 175
suertes, 119
susceptibilidad, 244, 247

T

tabaquismo, 176, 264, 266, 435
tablas
- cruzadas, 519
- de doble entrada, 519
- tetracóricas, 135, 190
- vacías, 459
tallos, 517
tamaño del efecto, 79, 80, 403
tamaño de muestra, 8, 63, 68-71, 73, 74, 76-89, 100, 106, 108, 111, 124, 126, 152, 155, 164, 166, 172, 181, 190, 202, 205, 207, 208, 218, 240, 252, 256, 262, 271, 283, 303, 313, 315, 319, 378, 416, 420, 455, 461, 493, 511, 515, 525, 541, 550, 565, 569

- casos y controles, 84
- cohortes, 85
- estudio transversal, 84
- modelos de regresión, 86, 87
- para evaluar hipótesis, 82
- - diferencia, 83
- para un promedio, 82
- para una proporción, 82
- pruebas diagnósticas, 86
tamaño del efecto, 77, 79, 303, 315, 317, 319, 400, 403, 548
tamizado, 19, 131, 132, 136, 170, 420, 421
tasa
- de descuento, 381-383, 403
- de incidencia, 115, 252, 253, 262
- - acumulada, 462
- de prevalencia simple, 461
- de verdaderos negativos, 135
- de verdaderos positivos, 134
técnica
- Delphi, 185
- de selección, 72
tecnología, 388
- de investigación clínica, 388
- combinadas, 388
- éticamente controversiales, 394
- obsoletas, 388
- sin controversias éticas, 394
tecnología *v.* evaluación económica, evaluaciones de tecnología
temperatura, 13
teoría de la decisión, 326
tercero pagador, 362
término MeSH, 33, 34, 42, 338, 474
test-retest, 109, 110
tiempo
- 0 (cero), 247
- de cohorte, 249
- de incidencia, 251
- de incubación, 249
- de latencia, 249
- hasta el evento, 251-253, 262
- libre de recaída, 251
- muertos, 157
toma de decisiones, 12, 19, 23, 32, 61, 147, 154, 286, 299, 300, 302, 306, 320, 325, 326, 330-332, 334-336, 341, 342, 349, 350, 352, 356, 358, 367, 376, 377, 386, 387, 389-392, 394, 400-402, 404, 406, 414, 431, 437, 455, 502, 531-534, 537, 538, 550, 551, 574, 588-590, 592
tradición sociológica, 290
trastorno
- bipolar, 238
- depresivo mayor, 102
- depresivo recurrente, 102
triangulación, 288
- de las fuentes de datos, 289
- de los investigadores, 289
- de los métodos, 289
trombosis venosa, 372
tuberculosis, 4, 228
- resistente, 434

U

úlcera, 123
- péptica, 123
umbral
- de costo-efectividad, 355
- de incertidumbre, 132, 133
unidad
- de análisis, 283
- de arranque, 70, 71
- de muestreo, 70, 72, 290
- muestral, 69, 72, 74
universo, 11, 45, 46, 335
- cardinal, 359

V

validación, 4, 92, 95-97, 99, 100, 102, 106, 108, 109, 156, 190, 258, 259, 288, 290, 374, 475-477, 479, 485-491
- de escalas, 92, 95, 97-111,
validez, 4-6, 12, 13, 16, 17, 33, 35, 36, 58, 59
- de constructo, 96, 108, 109
- de contenido, 102
- de criterio, 96, 106-108
- de los resultados, 77, 169, 234, 236, 352, 404, 406, 411, 445, 464, 465, 472, 566, 569, 577
- externa, 62, 75, 196, 233, 255, 257, 258, 337, 531, 532, 543, 544, 549, 550, 569, 573
- interna, 66, 75, 213, 233, 234, 255, 257, 258, 301, 532, 535, 543, 545, 566, 569
valor, 359
- atípico remoto, 515
- de *p,* 80, 104, 271, 535
- no fraccionario, 47
- predictivo negativo, 140, 143
- predictivo positivo, 140, 141, 143
variabilidad, 79
- interobservador, 216
- intraobservador, 216
variable, 46
- aleatorias, 46

- blandas, 443, 444
- categóricas, 82
- clasificación, 47
- control, 57
- definición, 46
- de confusión, 55, 193, 199, 205, 210, 216-219, 228, 232, 241, 242, 307, 372, 459, 462, 569
- de control, 57
- de desenlace, 204, 244
- de exposición de interés, 244
- de interés, 53, 547
- de tipo cuantitativo, 503
- dependiente, 198, 459
- diccionario de, 460
- duras, 443
- efecto de confusión, 462
- independiente, 86-88, 198, 457, 519
- intermediaria, 57
- medición, 48
- modificadoras del efecto, 199
- naturaleza, 47
- - discreta, 47, 50, 512
- - continua, 48, 264, 312, 460, 516, 519, 521
- no aleatorias, 46
- nominales, 507, 519
- predictoras, 53
- pronósticas, 199, 545
- relaciones, 53
variación
- aleatoria, 57, 59, 199, 548
- por el azar, 543
varianza a una vía, 276
varimax, 105
vasculitis, 172
ventana de interés, 374
veracidad, 5, 12, 13, 16, 17, 25, 26, 57, 58, 66, 421, 446, 572, 573, 583, 586
verdadero
- negativo, 126, 133, 135
- positivo, 133, 134
viabilidad, 8, 157, 556
virus
- de inmunodeficiencia humana (VIH), 136, 388, 404, 436
- del papiloma humano (VPH), 185
visitas de monitorización, 427, 428

Libros Medicos PDF

Álvaro J. Ruiz Morales • Carlos Gómez-Restrepo



Epidemiología Clínica

Investigación Clínica Aplicada

Este libro, dirigido a profesionales que trabajan en investigación con seres humanos en las áreas de la salud, se destaca por su enfoque práctico tanto en la apreciación crítica de la literatura como en la investigación clínica.

En sus páginas el lector encontrará explicaciones, recomendaciones y múltiples ejemplos acerca de cómo leer y evaluar los artículos científicos; cómo hacer investigación; cómo entender los aspectos básicos de la bioestadística, pero con un enfoque para clínicos; cómo aplicar los conceptos fundamentales de la epidemiología, sin ser epidemiólogo; cómo entender los estudios más relevantes en la actualidad, y cómo aplicarlos al día a día en la atención de pacientes, entre muchos otros temas. Además, esta nueva edición, revisada y ampliada, profundiza en aspectos tan importantes como la adaptación cultural de escalas, los diseños de investigación, el análisis de decisiones, la ética en epidemiología y los aspectos formales de presentación y autoría de resultados de investigación biomédica.

Sin duda, su contenido será muy útil para todo estudiante o profesional de la salud interesado en realizar investigación clínica o en evaluarla de forma crítica, así como para las facultades de medicina y los posgrados en medicina clínica de habla hispana y los comités de investigación con seres humanos.



EDITORIAL MEDICA panamericana