

MED

MEDI CINA

BASADA
EN LA

EVI
DENCIA

MEDICINA BASADA
EN LA EVIDENCIA
CONCEPTOS
INTRODUCTORIOS
PARA LA
PRÁCTICA CLÍNICA

Miguel Hernán Vicco
Luz María Rodeles Antonelli

UNIVERSIDAD
NACIONAL DEL LITORAL



Medicina Basada en la Evidencia



**UNIVERSIDAD
NACIONAL
DEL LITORAL**

Rector **Enrique Mammarella**
Director de Planeamiento y Gestión Académica **Daniel Comba**
Directora Ediciones UNL **Ivana Tosti**

.....

Vicco, Miguel Hernán
Medicina basada en la evidencia : conceptos
introdutorios para la práctica clínica / Miguel
Hernán Vicco ; Luz Rodeles ; contribuciones de
Héctor Musacchio ; Larisa Carrera ; Ponce Melisa;
prólogo de Oscar Bottasso. - 1a ed. - Santa Fe :
Ediciones UNL, 2020.
Libro digital, PDF - (Cátedra)

Archivo Digital: online
ISBN 978-987-749-180-7

2. Medicina Clínica. I. Rodeles, Luz. II. Musacchio,
Héctor, colab. III. Carrera, Larisa, colab. IV. Melisa,
Ponce, colab. V. Bottasso, Oscar, prolog. VI. Título.
CDD 614

.....

© Larisa Carrera, Héctor Musacchio,
Melisa Ponce, Luz María Rodeles Antonelli,
Miguel Hernán Vicco, 2020.

© del Prólogo, Oscar Bottasso, 2020.



© ediciones **UNL**, 2020

Coordinación editorial
María Alejandra Sedrán
Coordinación diseño
Alina Hill
Producción general
Ediciones UNL

—
editorial@unl.edu.ar
www.unl.edu.ar/editorial

.....



hdl.handle.net/11185/5517

Medicina Basada en la Evidencia

Conceptos introductorios para la práctica clínica

Miguel Hernán Vicco

Luz María Rodeles Antonelli

COORDINADORES

Héctor Musacchio

Larisa Carrera

Melisa Ponce

COLABORADORES



COLECCIÓN
CÁTEDRA

Índice

A propósito de prologar / 11

Prefacio / 15

1. EL CONCEPTO DE CIENCIA / 17

¿Qué quiere decir «ciencias» médicas? / 17

2. METODOLOGÍA DE LA INVESTIGACIÓN / 22

Concepto general / 22

3. MEDICINA BASADA EN LA EVIDENCIA / 25

¿Qué es la Medicina Basada en la Evidencia? / 25

4. EDUCACIÓN MÉDICA BASADA EN LA EVIDENCIA / 29

El rol de la MBE en la educación médica / 29

5. EL PASO A PASO DE LA MBE / 32

Etapas de la aplicación de la MBE / 32

Niveles de evidencia / 32

Etapas de aplicación / 35

6. NOCIONES BÁSICAS DE BIOESTADÍSTICA / 43

Conceptos generales / 43

Las pruebas estadísticas / 46

Asociación entre dos variables / 49

Validación de pruebas diagnósticas / 54

7. REPORTE DE CASO Y SERIE DE CASOS / 57

Generalidades / 57

8. ESTUDIOS TRANSVERSALES O TRANSACCIONALES / 59

Generalidades / 59

Análisis e interpretación de la asociación / 60

Limitantes de los trabajos transversales / 62

9. ESTUDIOS DE CASOS Y CONTROLES / 64

Generalidades / 64

Conformación de los grupos / 66

Análisis e interpretación de asociación / 67

Pautas para leer un trabajo casos y controles / 69

10. ESTUDIOS DE COHORTE	/ 71
Generalidades	/ 71
Conformación de los grupos	/ 73
Evaluación e interpretación de la asociación	/ 73
Pautas para leer un estudio de cohortes	/ 74
11. ANÁLISIS DE SOBREVIDA	/ 76
Generalidades	/ 76
Tiempo de seguimiento	/ 77
Métodos básicos de análisis de supervida	/ 78
Tabla de vida	/ 79
Curva de supervida	/ 81
12. ENSAYOS CLÍNICOS	/ 84
Generalidades	/ 84
Fases de un Ensayo Clínico	/ 85
Diseños de Ensayos Clínicos en relación con la asignación de los grupos	/ 86
Pautas para leer un Ensayo Clínico	/ 90
Evaluación e interpretación de los resultados	/ 93
13. ESTUDIOS DIAGNÓSTICOS	/ 96
Generalidades	/ 96
Pautas para la lectura crítica e interpretación de análisis estadísticos	/ 97
Resolución de la situación clínica	/ 103
14. METAANÁLISIS	/ 105
Generalidades	/ 105
Definición, ventajas y limitaciones	/ 106
Pautas para leer un metaanálisis	/ 107
Resolución de la situación problema	/ 112
15. FALACIAS EN MEDICINA	/ 114
Tangas, ojotas, Doritos, yoga y lentes de contacto (entre otras cosas)	/ 114
¿Asociación casual o causal?	/ 115
Comparaciones múltiples: la inflación no afecta solo a los precios	/ 116
Falacia de p	/ 117
NNT, NND y el riesgo relativo	/ 118
Punto final combinado	/ 119

Punto final sustituto / 119
Flexibilidad de los diseños y reporte selectivo de resultados / 121
Análisis de subgrupo *post-hoc* / 122
Estudios detenidos tempranamente: efectos «inflados» / 122
«Vibración del efecto»: cambio del resultado
 según el análisis estadístico / 123
Estudios de no inferioridad / 123
Sesgo de publicación / 124

REFERENCIAS BIBLIOGRÁFICAS / 125

SOBRE LOS AUTORES / 131

A mi familia y amigos

¿no estaba yo hacía ya muchísimo tiempo bastante alejado de la vida de todos los hombres, de la existencia y del pensamiento de las personas normales, no estaba yo hacía muchísimo tiempo bastante apartado y loco? Y, sin embargo, en lo más íntimo de mi ser comprendía perfectamente la llamada, la invitación a estar loco, a arrojar lejos de mí la razón, el obstáculo, a entregarme al mundo hondamente agitado y sin leyes del espíritu y de la fantasía.

El lobo estepario (Hermann Hesse)

A PROPÓSITO DE PROLOGAR

Obtener información sobre temas de relevancia en salud es un paso obligado de la práctica médica, sea para interiorizarnos sobre nuevos conocimientos fisiopatogénicos de una enfermedad o bien acerca de pautas diagnósticas y terapéuticas. Datos surgidos de la investigación clínica cuyo gran objetivo ha sido desde siempre tratar de reducir el nivel de incertidumbre.

Sin llegar a constituir una certeza absoluta, en muchas situaciones el médico transita sobre un terreno bien afirmado y siente que está adoptando decisiones sobre la base de criterios perfectamente fundados, lo cual otorga seguridad a su accionar. Pero más temprano que tarde se harán presentes los nubarrones, algunos bastante amenazantes, por cierto. Es en ese preciso momento cuando emerge la pregunta tantísimas veces formulada: ¿qué se sabe de esto, alguien conoce de estudios serios al respecto?

Hacer frente a las incertidumbres en torno a determinado problema médico en el contexto de un sistema de salud y la singularidad de cada paciente ha venido siendo el gran mandato de la investigación médica. Afortunadamente, y gracias a la labor colectiva de tantos equipos en el mundo, hoy contamos con técnicas de análisis en cuanto a cómo manejarnos ante cuestiones irresolutas del modo más coherente posible, basados en inferencias razonables y esperanzados en que mañana podríamos contar con datos superadores para despejar aún más las incertezas del presente. Como nunca antes, asistimos a un tiempo caracterizado por un aumento exponencial de trabajos de investigación, casi un millón de nuevos estudios por año en el campo de la salud. Este número creciente de publicaciones científicas requiere un sistema de evaluación lo suficientemente crítico para ponderar la calidad, pertinencia y aplicabilidad de

los resultados aportados y finalmente dirimir si ello es trasladable a la práctica clínica. En definitiva, necesitamos obtener datos confiables para así sentirnos más serenos ante la complejidad.

San Agustín, en sus *Confesiones*, hace una reflexión que puede aplicarse al caso que nos ocupa: «era un tiempo en que no había alcanzado la verdad, pero sí había aprendido a identificar las falacias». Lamentablemente, en el campo médico no disponemos de libros sapienciales ni tampoco de profetas capaces de vaticinar el devenir de tal o cual padecimiento, más allá de que alguna vez nos hayamos topado con algún autodeclarado gurú, pero ese mal trance no debe endilgarse a la ciencia sino a la estupidez humana. Lo que sí podemos hacer es diferenciar los estudios valiosos de aquellos que no lo son y de ese modo tomar el camino menos desatinado. Claro está que para llevar a cabo tales discernimientos debemos tener conocimiento sobre las bondades y limitaciones de cada tipo de investigación en particular, lo cual nos lleva a trasladarnos a la arena de la Metodología y la Medicina Basada en la Evidencia (MBE).

El texto de los colegas Vicco y colaboradores es una contribución tan valiosa como necesaria en este terreno y seguramente hará posible que el lector consiga adentrarse en las principales estrategias que utiliza la investigación clínica para el abordaje de problemas médicos como también las fortalezas y limitaciones de cada tipo de diseño. Los diferentes capítulos en que se desagrega el libro están engarzados en una suerte de escalado que facilita aún más el entendimiento de la lógica en la que se mueve la labor investigativa. En líneas generales, el aspecto formativo del estudiante de medicina en el campo de la investigación científica no pasa de ser un vistazo con un variable grado de detenimiento según los diferentes escenarios. Disponer de un texto de Metodología es siempre auspicioso y más cuando sus redactores, imbuidos ya en las artes del hurgar prolijamente, están a la vuelta de la esquina y con oídos atentos ante el entusiasta deseo de ingresar a la cofradía.

La tarea de interiorizar al estudiante o graduado sobre los pormenores que hacen a la MBE también halla su justificación habida cuenta de los argumentos esgrimidos en su contra; entre los más ventilados, el acotamiento de la libertad del médico. Pero hasta qué punto podemos llegar a ser realmente libres cuando somos aturcidos por nuevas ofertas de herramientas diagnósticas y potenciales estrategias terapéuticas de un costo cada vez más

creciente y preocupante para cualquier país. Ningún sistema de salud tendría recursos suficientes para proveer a todos sus ciudadanos de procedimientos onerosos cuyo balance entre efectividad y costo no siempre guarda una relación civilizada. De no contar con abordajes que permitieran una correcta justipreciación de la oferta podríamos caer en una confusión no exenta de riesgos. Esto fue en gran medida la fuerza impulsora de la MBE como uso consciente, claro y sensato de la mejor evidencia a la hora de tomar decisiones sobre el cuidado de un paciente. En un escenario de negociaciones donde, por un lado, se ubica el profesional, idóneo y consecuentemente capaz de ponderar el peso de la situación clínica y los datos de la literatura y, por el otro, el paciente con sus expectativas y que en función de su autonomía puede aspirar a tener un rol activo respecto de las decisiones que se adoptarán sobre su persona. Un enfermo que hoy visita sitios de la web para informarse sobre su dolencia y posteriormente interpelar al médico con cuestiones no siempre sencillas de zanjar, lo que suma otro ingrediente para seguir consultando las bases de datos. Con los debidos recaudos, a fin de resguardarnos de aquellos que han aprendido demasiado siempre ávidos por ocupar espacios de opinión en un terreno donde cualquier intento de dogmatismo en nombre de la MBE es una flagrante contradicción.

Desde su misma base epistemológica, el conocimiento surgido de estudios clínicos incuestionables no deja de ser una proposición temporalmente aceptada. Muy satisfechos, no obstante, porque haber arribado a ello no es poca cosa, pero sin perder de vista aquella vieja sentencia de los autores anglosajones: *the best is to come*.

La problemática del conocimiento científico es un hecho no lo suficientemente discutido y divulgado en nuestras facultades de Medicina, por lo cual el libro *Medicina Basada en la Evidencia* es a todas luces bienvenido. Mis plácemes para Luz y Miguel, dos colegas que, además de sus capacidades, poseen la pasión indispensable para llevar adelante la actividad insoslayable de la formación universitaria.

Dr. Oscar Bottasso
Instituto de Inmunología
Clínica y Experimental de Rosario
UNR-CONICET

PREFACIO

Desde el inicio de la praxis médica, la búsqueda de las causas de las enfermedades y las modalidades de abordar las mismas para brindar bienestar a los pacientes fueron motores del continuo desarrollo de saberes. Estos, por suerte, siempre estuvieron sujetos al cuestionamiento y comprobación por parte de alguien no saciado con una simple respuesta. Las mentes inquietas, aquellos que nunca tomaron como verdades absolutas las enseñanzas de sus maestros, fueron las culpables de que en el transcurso de la historia de la medicina se sucedieran los cambios de enfoques del proceso salud-enfermedad. A ellos les debemos que el concepto de enfermedad deje de ser «castigo divino» y sea considerado como el desequilibrio bio-psico-social, el auge de las ciencias biológicas, como la fisiología, la histología o la anatomía patológica, y la influencia del contexto social en los procesos patológicos.

A fines del siglo XIX se produce el mayor crecimiento tecno-científico en el campo de la medicina, que en el contexto del Estado social promueve el proceso de medicalización social:

Con mucha frecuencia la medicina se impone al individuo, enfermo o no, como un acto de autoridad. A este respecto pueden citarse varios ejemplos. En la actualidad no se contrata a nadie sin el dictamen del médico que examina autoritariamente al individuo. Existe una política sistemática y obligatoria de localización de enfermedades en la población que no responde a ninguna demanda del enfermo. (Foucault, 1994)

A partir de este, los cuestionamientos dejan de estar abocados a la veracidad del saber y se expanden hacia la incorporación y práctica de las innovaciones tecno-científicas en el proceso salud-enfermedad. «¿Es necesario realizar este o aquel estudio?

¿Vale la pena realizar screening de una patología, de cuál? ¿Qué tratamiento se debe aplicar, es efectivo?».

Recién en la década del 90, David Sackett bautiza este accionar como «Medicina Basada en la Evidencia» (MBE), y lo define como uso consciente, explícito y juicioso de la mejor evidencia actual en la toma de decisiones sobre el cuidado de cada paciente. Existen detractores de esta corriente que argumentan que su práctica despersonaliza al médico y la relación que establece con el paciente. Sin embargo, MBE es la promoción de un saber reflexivo. Mañana, tal vez, MBE se denomine de otra forma, y no faltará quien diga que dejó de existir, pero el escepticismo del saber médico va a continuar porque aún hay preguntas sin respuestas y porque no hay verdades absolutas.

A lo largo de este libro intentaremos acercar conceptos básicos necesarios para el análisis de los problemas desde la perspectiva de la MBE. El objetivo es hacer un aporte al proceso que el estudiante de ciencias médicas transitará para avanzar en la construcción de un razonamiento crítico. La praxis médica necesita de la búsqueda eficiente y del análisis pormenorizado de los artículos científicos como herramienta de enriquecimiento, de actualización y de decisión clínica. Requiere, asimismo, de la duda sana y constante, un espíritu dispuesto a navegar también en los inmensos espacios de la incertidumbre. Desde la MBE se nos proponen algunos instrumentos que compendiamos brevemente en estas páginas, para asistirnos en la dificultosa tarea de intentar guiar a otro ser que acude a nosotros, que transita nada menos que su dolencia.

Por este motivo, agradecemos la posibilidad de contribuir en estos primeros pasos y esperamos que sean de utilidad. Invitamos a que cada uno pueda acercarnos sus preguntas o aportar aquellas cuestiones que considere que necesiten más claridad, ya que esta interacción es un objetivo esencial y constitutivo del aprendizaje conjunto que les proponemos emprender.

1 El concepto de ciencia

¿QUÉ QUIERE DECIR «CIENCIAS» MÉDICAS?

Sin pretender realizar una revisión epistemológica, cuya riqueza y extensión escapan a lo propuesto por esta presentación, cabe comenzar por retomar algunas ideas sobre el concepto de «ciencia» que nos ayuden a comprender cómo la cuestión derivó, en el campo médico, en la Medicina Basada en la Evidencia (MBE).

¿Qué quiere decir «ciencia»? Desde su origen, el hombre ha construido (y aplicado) el saber empírico desde la experiencia social y de las relaciones humanas. Sin embargo, en el momento en que comienza a reflexionar sobre el saber y trata de establecerlo como verdad dicha por una autoridad, encuentra dificultad para justificar ese conocimiento como «ley del saber». Entonces surge, en la antigua Grecia, la investigación del origen, certeza y extensión del conocimiento humano junto con sus fundamentos y grado de creencia, opinión y asentimientos. Pero no es hasta el siglo XVII que el conocimiento científico comienza a adquirir mayor impacto en el desarrollo del saber, y ese es el origen de la ciencia moderna. Se inicia un proceso en el cual los sabios cuestionan las verdades absolutas y procuran obtener nuevos conocimientos mediante el método experimental, abandonan el principio de autoridad que determinaba que las cosas eran verdades según quién lo decía.

El sociólogo Robert Merton postulaba que ciencia es un concepto utilizado con el fin de denotar al conjunto de métodos por medio de los cuales se certifica el saber. Buscaba explicar el surgimiento de la ciencia mediante la diferenciación entre el conjunto de conocimiento empírico y aquel como resultado de la experimentación. Un ejemplo de esto fue la controversia con la que se encontró Claude Bernard en el siglo XIX. Por aquel entonces el saber empírico decía que el tinte rojo de la sangre era consecuencia de la presencia de oxígeno (por lo que, si no había oxígeno, la sangre no podía ser roja). Sin embargo, Bernard, luego de 8 años de realizar experimentos, pudo demostrar que esto no era una verdad absoluta, ya que el color de la sangre con carboxihemoglobina era similar al que contenía oxihemoglobina. El saber empírico quedó de lado, el científico comenzó a desplazar las verdades absolutas.

Además de provenir de la experimentación, el saber científico debe de cumplir con ciertas normas, que son:

- Universalismo: los conocimientos deben ser sometidos a criterios impersonales preestablecidos.
- Desinterés: control institucional que impone la institución científica y ante la cual los científicos se someten. Los resultados de los científicos deben ser puestos a consideración en un proceso de verificación por sus pares, y este control «policíaco» debe velar por garantizar, por ejemplo, la casi inexistencia del fraude científico.
- Escepticismo organizado: el científico tiene que someter todo a un análisis crítico; para él, nada debe ser sagrado, no debe tener prejuicios.

Un rasgo del conocimiento científico que deseo subrayar es su generalidad. Si consideramos ejemplos no problemáticos de conocimiento científico, por ejemplo la geometría euclídeana y la ley de reflexión de luz, conocidas por los antiguos, o la mecánica de Newton y la teoría de la relatividad de Einstein en tiempos más modernos, no es difícil apreciar la generalidad de las afirmaciones entrañadas. Los teoremas de la geometría se aplican por igual en los dominios de la carpintería, la agrimensura y la astronomía, mientras que la mecánica newtoniana se aplica tanto al movimiento de los cometas como a la oscilación de un péndulo. (Chalmers, 1990)

Con independencia de sus cualidades, la ciencia es producto de la interacción de personas que buscan concretar un fin común, plasmado en el concepto de colectividad científica descrito por Michael Polanyi, o de campo científico propuesto por Pierre Bourdieu, por cuanto implica un espacio de conflicto, de lucha, mediante el cual surge el conocimiento científico.

El campo científico, como sistema de relaciones objetivas entre posiciones adquiridas (en las luchas anteriores) es el lugar (es decir el espacio de juego) de una lucha competitiva que tiene por desafío específico el monopolio de la autoridad científica, inseparablemente definida como capacidad técnica y como poder social, o si se prefiere, el monopolio de la competencia científica que es socialmente reconocida a un agente determinado, entendida en el sentido de capacidad de hablar e intervenir legítimamente (es decir, de manera autorizada y con autoridad) en materia de ciencia. (Bourdieu, 1976)

En contraposición al campo científico, Karin Knorr-Cetina propone el concepto de arenas transepistémicas de investigación. Refiere al espacio donde se desempeñan los científicos y se interrelacionan y a los recursos utilizados. Quiere significar, por un lado, que el espacio en el que se desenvuelven los investigadores no es fijo sino que se modifica conforme avanzan las investigaciones de los mismos, pues sus relaciones de recursos se van redefiniendo; por el otro, busca resaltar que los elementos que intervienen en el proceso son heterogéneos y no son propios del saber científico.

En este contexto, se aprecian los vínculos entre los aspectos cognitivos y los sociales que lo condicionan y se abre así la caja negra del conocimiento científico con Thomas Kuhn. Esto se debe a que, a partir de él, las categorías de paradigma, ciencia normal y las revoluciones científicas, permiten explorar la producción misma del conocimiento científico.

La ciencia normal es la que produce los ladrillos que la investigación científica está continuamente añadiendo al creciente edificio del conocimiento científico. Esta concepción acumulativa del desarrollo científico es familiar y ha guiado la elaboración de una considerable literatura metodológica. Tanto esta concepción como sus subproductos metodológicos se aplican a una gran cantidad de trabajo científico importante. Pero el desarrollo científico manifiesta también una modalidad no acumulativa, y los episodios que la exhiben proporcionan claves únicas de un aspecto central del conocimiento científico. (Kuhn, 1990)

Según este pensador, los científicos determinan la significancia de un hecho mediante su articulación con la teoría en el proceso de la ciencia normal manteniendo coherencia con el paradigma que prima en el momento. Durante las crisis revolucionarias surge la confusión y se carece de postulados firmes, y es entonces cuando intervienen factores extracientíficos para incidir en una toma de decisión. Durante estos períodos no se puede optar entre distintas teorías solo apelando a cuestiones cognitivas: la explicación sociológica se vuelve fundamental e irreductible, se modifica la concepción del mundo con el cambio del paradigma.

El conocimiento científico es el resultado de una construcción social, razón por la cual se la enmarca en el «constructivismo». La ciencia toma el carácter de actividad social. Los científicos no son engranajes que aplican un método sino que interactúan en distintos planos para concretar el objetivo científico. David Bloor es un interesante representante y uno de los grandes propulsores de la corriente constructivista; conforma el programa de estudios sobre la ciencia denominado «Programa Fuerte», que se sostiene sobre cuatro principios muy importantes:

- 1) *Causalidad*: debe ser causal, es decir, ocuparse de las condiciones que dan lugar a las creencias o a los estados de conocimiento. Naturalmente, habrá otros tipos de causas, además de las sociales, que contribuyan a dar lugar a una creencia.
- 2) *Imparcialidad*: tiene que ser imparcial con respecto a la verdad y la falsedad, la racionalidad y la irracionalidad, el éxito y el fracaso. Y estas dicotomías exigen explicación.
- 3) *Simetría*: es necesario que sea simétrica en su estilo de explicación. Los mismos tipos de causas deben explicar, digamos, las creencias falsas y las verdaderas.
- 4) *Reflexibilidad*: debe ser reflexiva. En principio, sus patrones de explicación deberían ser aplicables a la sociología misma. Como el requisito

de simetría, éste es una respuesta a la necesidad de buscar explicaciones generales. Se trata de un requerimiento obvio de principio porque, de otro modo, la sociología sería una refutación viva de sus propias teorías.

Harry Collins, por su parte, rige sus investigaciones en lo social basándose en los principios de imparcialidad y simetría, vierte una mirada crítica sobre el Programa Fuerte y desarrolla su propio abordaje, al que denomina «Programa Empírico del Relativismo». Collins, al desarrollar el principio de simetría, cuestiona las explicaciones sobre el conocimiento que recurren a lo verdadero, racional, exitoso y progresivo (*True, Rational, Successful, Progressive* – TRASP) como fundamento último de la explicación. Advierte que ninguno de los conceptos TRASP puede servir para explicar la producción de conocimiento porque no permiten explicar por igual todos los trabajos científicos. Quienes están participando de un experimento no tienen ninguna certeza de si habrán de producir un conocimiento verdadero o exitoso. Ocurre, incluso, que en las prácticas científicas pueden involucrarse numerosos elementos de irracionalidad. Sin embargo, cuando un científico revisa su propia trayectoria le imprime una racionalidad a sus prácticas que seguramente no tenían en su momento. Lo que hoy aparece como un resultado prometedor, en su momento era una mezcla de confusas prácticas y discursos.

Con Harry Collins surge el estudio de las controversias científicas, escenario ideal para analizar no solo la producción científica sino los mecanismos mediante los cuales se entrelazan científicos en la refutación o reafirmación de los hechos. Los datos, en sí mismos, no resuelven nada, pues están sujetos a una flexibilidad interpretativa. En el análisis de las controversias es posible observar cómo los actores concretos llegan finalmente a una clausura, instaurándose un consenso sobre cuál es el hecho científico y cuál no. Resalta la importancia de analizar a los grupos particulares que toman parte de la controversia y señala que existe un núcleo central (*core set*), constituido por los actores principales, los que están en el centro de las discusiones y realizan los aportes más significativos. Para entender cómo se cierra una controversia es preciso estudiar estos núcleos centrales, seguir su trayectoria, ver cómo se comportan.

Es evidente la presencia, en estos abordajes, de ciertas ideas de Kuhn. Hay una clara reminiscencia a la idea de paradigma, del consenso que implica, y de las crisis de la ciencia normal. Pero estas categorías aparecen reformuladas ahora en un marco de análisis muy concreto. De hecho, es importante entender el motivo por el cual las controversias científicas son un espacio tan valioso, simplemente porque ellas se tratan precisamente de las anomalías de la ciencia normal. Cuando los científicos producen resultados que no cuadran bien en un paradigma e intentan producir algunos cambios en la estructura de este, se ven envueltos en una controversia.

Al afirmar que la medicina es una ciencia, reconociendo que su objeto de estudio es la salud humana, queda sujeta a su historia para comprender su configuración, su práctica y su inserción en la sociedad como elemento de saber, pues «si se considera a la historia como algo más que un depósito de anécdotas o cronología, puede producirse una transformación decisiva de la concepción que tenemos de la ciencia» (Kuhn, 1990). Desde la posición de Georges Canguilhem existen dos enfoques desde los cuales se puede analizar el desarrollo de la ciencia, uno es el «internalista», para el cual el contexto no interviene en el mismo; mientras que el otro, «externalista», considera que las acciones médicas se encuentran sujetas a la interacción del hombre con su medio, a las situaciones que constituyen el todo social.

Desde esta última posición, para lograr una correcta reconstrucción del crecimiento científico se deben tener en cuenta los discursos sociopolíticos y las relaciones de poder que lo configuran. Tal como refiere Álvaro Cardona

las condiciones de salud y bienestar de los colectivos poblacionales en cada circunstancia histórico social específica, que en razón de su historicidad se transforman permanentemente, siguiendo alguna lógica más o menos sistemática cuyo desentrañamiento reta el esfuerzo científico de la disciplina. (2008)

Entonces, con respecto a la medicina en su dimensión científica, da valor a sus afirmaciones y teorías en función de los procedimientos experimentales y estadísticos. Se fundamenta en leyes de generalización empírica y formula pronósticos que no serán otra cosa que hipótesis médicas, es decir, enunciados contrastables para su confirmación o refutación. Pero, sin duda, no podemos abarcar la medicina solo con estas cuestiones.

Edmundo Pellegrino (2001) menciona que «la medicina es la más humana de las ciencias, la más empírica de las artes y la más científica de las humanidades»; entonces hablar de ciencias médicas no es hablar de un proceso aislado que solo satisface a quien lo conforma, o el resultado de un enlace de métodos estandarizados en busca de cumplir con un objetivo. Es la puesta en práctica del entendimiento de la naturaleza que reúne a la medicina, la comprensión del proceso salud–persona–enfermedad en la satisfacción de las necesidades sociales.

2 Metodología de la investigación

CONCEPTO GENERAL

Referimos en el capítulo anterior que el término «ciencia» se emplea con el objetivo de clasificar a las disciplinas que hacen uso de las formas por medio de las cuales certifican el conocimiento. Parece simple, pero no lo es. Ciencia implica investigación, la búsqueda sistematizada de información sobre un tópico específico. Es entonces cuando se introduce el concepto de «metodología», es decir, el estudio o tratado de los procedimientos con el fin de encontrar soluciones a diversas problemáticas.

En pocas palabras, metodología de la investigación es el proceso para desarrollar de manera eficaz la comprobación del saber. Los aspectos generales para realizar la investigación comprenden:

- a) Planteo de las hipótesis: se trata del supuesto teórico que se comprobará mediante el desarrollo de la investigación. Su expresión debe ser clara, delimitar el objeto de estudio (parte de la realidad que se va a investigar) y qué es lo que se quiere valorar (objetivos). ¿Qué se investiga?
- b) Conformación del marco teórico: establece, mediante revisión de antecedentes sobre el problema en cuestión, la contextualización y justificación teórica de la hipótesis. ¿Por qué se supone que la hipótesis podría ser real?
- c) Prueba de hipótesis: en esta etapa se seleccionan las herramientas que se emplearán para analizar las variables que intervienen en los objetivos planteados anteriormente y conforman nuestra hipótesis. Se pauta la metodología a emplear. ¿Cómo se va a estudiar la hipótesis?
- d) Resultados y conclusión: luego de pautar las pruebas a realizar para contrastar la hipótesis, se lleva adelante la recolección y análisis de datos. Aquí podremos confirmar o no el planteo teórico. ¿Cuáles fueron las observaciones del contraste de la hipótesis?

Estas etapas se pueden esquematizar de la siguiente forma:

TABLA 1. ESQUEMA DE LAS ETAPAS DE INVESTIGACIÓN

Coherencia Interna				
Hipótesis y Marco Teórico	Objetivo General	Objetivos Específicos	Metodología	Resultados / Conclusión

Si observamos un artículo científico, podremos ver que su estructura es similar al cuadro anterior. Para ejemplificar lo antes expuesto, procedamos a identificar las etapas en el resumen del artículo «Alteraciones metabólicas en hijos de padres con hipertensión arterial esencial» publicado en la revista *Medicina*, 73(3) (2013:243–246):

- Hipótesis, Marco teórico y Objetivo General: el antecedente familiar de hipertensión arterial en jóvenes sanos se ha asociado a hiperinsulinemia, que a su vez produciría aumento en el cortisol sérico, confluyendo ambos mecanismos en daño endotelial renal con la presencia de microalbuminuria. El objetivo del estudio consistió en evaluar en jóvenes sanos, hijos de hipertensos, la asociación entre los niveles de insulinemia, cortisol sérico y microalbuminuria, debido a su relación con mayor riesgo cardiovascular.
- Metodología: se realizó un trabajo transeccional y correlacional en la ciudad de Santa Fe con 145 jóvenes sanos mayores de 18 años de edad que se asignaron a dos grupos: uno con los que tenían antecedente de primer grado de hipertensión arterial esencial (grupo de estudio) y otro con los que no contaban con dicho antecedente (grupo control). Se valoraron las concentraciones séricas en ayunas de insulina, cortisol, y los niveles de microalbuminuria en primera orina matutina.
- Resultados y conclusión: la media de edad fue de $20 \pm 2,9$ años, de lo cual en un 58 % se trató de mujeres. El grupo de estudio incluyó el 48 % ($n= 69$). El 4,8 % presentó insulinoresistencia, 13,8 % microalbuminuria, y el 52 % hipercortisolínemia, no encontrándose diferencias significativas de los niveles séricos de insulina y cortisol ni de microalbuminuria entre los grupos, como tampoco correlación entre estas variables. No se encontró asociación entre el antecedente de 1º grado de hipertensión arterial y alteraciones de la homeostasis de insulina o cortisol y tampoco evidencia de daño endotelial con presencia de microalbuminuria.

Dentro de las ciencias médicas, habitualmente los trabajos científicos pueden clasificarse en descriptivos, analíticos y experimentales. El primero de ellos es el único que no requiere de una hipótesis, ya que mediante el mismo solo se detallan características o cualidades particulares de las variables observadas.

Un ejemplo de este tipo de investigación son los reportes de casos o series de caso. Por su parte, los trabajos analíticos y experimentales implican el estudio de las variables para contrastar la hipótesis planteada. Se diferencian entre ellos en que en el caso de las investigaciones experimentales el investigador interviene introduciendo un factor (variable de interés) para ser estudiado.

Existen dos grandes grupos de investigaciones que se pueden realizar en forma independiente o complementándose entre ellos:

- a) Cualitativas: los datos recolectados no se cuantifican sino que traducen la apreciación u opiniones dada por el investigador o por personas entrevistadas con relación a un aspecto particular (objeto de estudio). Por ejemplo: ¿cuál es su opinión sobre la eutanasia?
- b) Cuantitativas: este tipo de investigación implica la medición a través de valores objetivos. Por ejemplo: ¿cuál es el valor de la glucemia en ayunas?

Existe amplia bibliografía que compara la calidad entre una u otra forma de realizar la investigación. En particular, consideramos que no hay una mejor que otra, son formas distintas de abordar diferentes tipos de problemas respecto de cómo se plantea su análisis y la naturaleza de su objeto de estudio. En el caso puntual de la medicina, así como en la generalidad de las ciencias biomédicas, la gran mayoría de la bibliografía responde a investigaciones cuantitativas, por lo cual nos centraremos en ellas.

La práctica diaria de medicina como disciplina científica se encuentra sujeta al continuo al contraste de los supuestos teóricos, sea para determinar factores de riesgo, para valorar pruebas diagnósticas o eficacia de nuevos fármacos. La investigación es una actividad fundamental dentro del campo médico que, además de aportar sus observaciones, obliga al profesional a mantenerse actualizado de manera constante. No nos olvidemos de que si la medicina es considerada ciencia debe cumplir con los preceptos de universalidad, desinterés y escepticismo. No existen verdades absolutas ni grandes profetas, por el contrario, como refería el educador médico Sydney Burwell (Past Dean of the Faculty of Medicine, 1935–1949, Harvard Medical School): *«half of what we are going to teach you is wrong, and half of it is right. Our problem is that we don't know which half is which»*. Ciencia implica ratificar o rectificar las observaciones comunicadas por los diferentes grupos de investigación. Podemos decir que 10 años es hoy en día mucho tiempo puesto que los cambios de los supuestos teóricos cada vez se suceden más rápido.

3 Medicina Basada en la Evidencia

La ignorancia es el saber erróneo o no saber. Si desconocemos el hecho de que los gérmenes causan enfermedades, somos ignorantes. Este desconocimiento en un médico del siglo XXI tendría un grave desenlace, pero no en un profesional del siglo IV. En ese entonces la ignorancia era colectiva, no se sabía que los microorganismos podían causar enfermedades. Hoy se sabe, por lo que desconocerlo hace al ignorante individual que pone en riesgo a sus pacientes.
(Kenneth Goodman)

¿QUÉ ES LA MEDICINA BASADA EN LA EVIDENCIA?

En el transcurso de la década del 90, el médico epidemiólogo David Sackett promulga el concepto de Medicina Basada en la Evidencia (MBE) definiéndolo como el uso consciente, explícito y juicioso de la mejor evidencia actual en la toma de decisiones sobre el cuidado de cada paciente, con el objetivo evaluar la calidad de la evidencia respecto de los riesgos y beneficios tanto de la aplicación de los distintos estudios complementarios en la valoración de un diagnóstico presuntivo como del tratamiento. En pocas palabras, la MBE se incorpora al ejercicio profesional como un puente entre la experiencia y el desarrollo científico.

Durante la implementación de los distintos métodos para lograr las metas de la MBE interviene la valoración de la tecnología sanitaria, definida por la Sociedad Internacional de Farmacoeconomía y Resultados de Investigación como política de investigación que examina consecuencias a corto y largo plazo de la aplicación de una tecnología de atención de salud. Las propiedades evaluadas incluyen pruebas de seguridad, eficacia estadística, resultados de las respuestas en pacientes, eficacia clínica, el costo y la rentabilidad, así como impactos sociales, jurídicos, éticos y políticos. Este análisis es de metodología lineal y consecutiva, ya que uno no puede estimar la magnitud de los resultados clínicos sin antes evaluar las pruebas clínicas. No se pueden comparar los costos y rentabilidad sin estimar los efectos sobre los

resultados clínicos; y no se puede pensar sobre las implicaciones éticas y jurídicas de las tecnologías hasta que uno sabe algo acerca de los costos y la rentabilidad. Este proceso comienza con la evaluación de la evidencia, y en este sentido la MBE es parte de la valoración tecnológica.

La MBE involucra dos aspectos: por un lado, la educación del profesional de salud en la aplicación del conocimiento científico orientando la toma de decisiones sustentadas en la mejor evidencia, sintetizadas mediante el juicio clínico y enfocadas en el paciente; por otro lado, se encuentran las guías basadas en evidencia o políticas basadas en evidencia. Esta parte describe la importancia de políticas fundadas en la población, políticas de cobertura y valoración del rendimiento de la evidencia; hace hincapié en el principio de que, antes de promover políticas poblacionales, debe haber evidencia suficiente de que sea efectiva y beneficiosa.

Se puede esquematizar la práctica de la MBE en la valoración de la innovación científico-tecnológica de la siguiente forma:

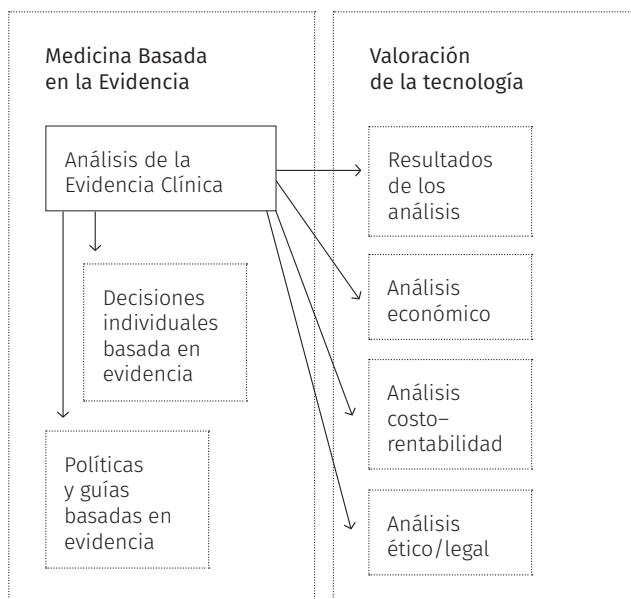


FIGURA 1. ESQUEMA DE LA PRÁCTICA DE LA MBE

El objetivo principal de la MBE es reducir la brecha entre la ciencia y la tecnología y el conocimiento práctico del profesional al permitir que las decisiones médicas se basen en la mejor evidencia científica. Para lograr este fin, el proceso implica cuatro pasos:

- 1) Hacer de un dilema médico una pregunta que pueda tener respuesta.
- 2) Buscar la mejor evidencia.

- 3) Evaluar críticamente la evidencia.
- 4) Aplicar la información obtenida en la práctica clínica.

Sobre este punto surgen las críticas o cuestionamientos a este modelo, como por ejemplo, ¿cuál es la mejor evidencia? ¿Por qué la estadística está sobre la experiencia? O, como se planteaba Richard Horton, editor de la revista científica *The Lancet*: ¿y el paciente?

En tanto, frente a la cuestión: ¿cuál es el mejor tratamiento para el paciente medio estandarizado?, surge esta otra pregunta: ¿cuál es el mejor tratamiento para el paciente que tengo sentado a mi lado? Tratar de convertir la enfermedad en objetiva es negar la primacía del enfermo en la medicina.

Todos estos planteos son válidos, pero: ¿por qué surgen ahora, más precisamente a partir de la década del 90 y no antes? A *prima facie*, uno podría replicar: porque la medicina basada en evidencia es de ahora. Y es verdad, este concepto surgió hace poco, pero no necesariamente implica algo que se aplica recientemente. ¿Cuántos realizan sangrías para el tratamiento de neumonías? ¿Quién puede negar la utilidad del cardiodesfibrilador en la reanimación cardíaca? La respuesta es simple y nos parece hasta obvia: nadie. Aceptamos como cotidiano el uso del desfibrilador, y aplicar sangría en un paciente con neumonía nos llevaría a pecar de ignorantes, pues ambos descubrimientos fueron demostrados mediante ensayos clínicos durante el siglo XIX, cien años antes que el Dr. David Sackett mencionara por primera vez MBE.

Pierre Louis, además de criticar la práctica habitual de sangría para el tratamiento de neumonías o gripes durante el siglo XIX con la realización del ensayo clínico para desestimar su uso, introdujo el método numérico para lograr la «comprobación» científica del ejercicio.

As to different methods of treatment, it is possible for us to assure ourselves of the superiority of one or other (...). by enquiring if the greater number of individuals have been cured by one means than another. Here it is necessary to count. And it is, in great part at least, because hitherto this method has not at all, or rarely been employed, that the science of therapeutics is so uncertain. (1834)

El objetivo perseguido a través de este método era evitar tratamientos cuya eficacia no estuviera realmente demostrada.

El ejercicio de la medicina estuvo, a lo largo de su historia, basado en evidencia desarrollada desde las distintas concepciones que fue adquiriendo el binomio salud–enfermedad. Siempre existieron nuevos conceptos desafiantes para las enseñanzas de maestros, pero hoy nos confrontamos a un obstáculo epistemológico que va más allá de la MBE propuesta por el Dr. David Sackett. La producción de conocimiento científico es casi vertiginosa y está al alcance de la mayoría de los médicos. Ya no es una sociedad científica la que transmite los avances en medicina, sino que día a día hay miles de *papers* que

aportan algo a nuestro saber y casi a la vuelta de la esquina. Más aún, no solo puede acceder el médico sino los mismos pacientes que quieren adentrarse en el conocimiento de su padecimiento. Así las cosas, desconocer el *paper* publicado en tal o cual *journal* nos coloca en una posición hartamente incómoda.

Esta necesidad de actualizarnos en forma continua sobre el binomio salud-enfermedad nos obliga a saber reconocer la calidad de un artículo, su validez interna y externa, nociones básicas de metodología para poder interpretar los resultados, el nivel de la revista en la que se publica, quién lo escribe (y los intereses comerciales de trasfondo), y si realmente nos sirve para respondernos el interrogante que surge en la consulta con nuestro paciente.

No es la MBE un sistema ajeno a la práctica cotidiana pues implicaría que el día a día del médico no se sustenta en un saber científico, y eso hace cuestionable nuestra labor en beneficio del paciente. Por el contrario, es la demanda de la oferta cuasi infinita de conocimiento la que nos pone continuamente ante el desafío de pecar de «ignorantes individuales».

4 Educación médica basada en la evidencia

No se puede caer en la posición que centra en los profesores toda la responsabilidad a la hora de mejorar la enseñanza, pero sí nos parece justo admitir que ningún cambio en ella puede hacerse sin su participación, sin su transformación.
(Gimeno Sacristán)

EL ROL DE LA MBE EN LA EDUCACIÓN MÉDICA

Este libro muestra en sus diferentes capítulos que, en medicina, la toma de decisiones debe fundarse en estudios válidos y fiables de investigación, cuyos resultados y conclusiones avalen las decisiones a adoptar frente a cada paciente. Ello es así porque, entre otras razones, el aumento de la disponibilidad de tecnología sanitaria, el incremento de los costos y gastos sanitarios y la cantidad de información, cada vez más importante, constituyen verdaderos desafíos para la práctica clínica y determinan la creciente necesidad de información confiable que fundamenten tales decisiones. El propio sistema de salud, con sus recursos limitados, debe lograr identificar los procedimientos que garanticen al máximo su utilización en beneficio de la población y de la calidad y racionalidad de estas prestaciones.

Esto, que resulta palmario con relación a la praxis médica, lleva a preguntar qué sucede en lo que atañe a la educación médica.

Formar parte de una institución educativa implica recurrir muchas veces a toma de decisiones educativas que requieren del convencimiento de quienes se decantan por ellas. Empero, ello también incluye a los equipos docentes que se verán involucrados en los cambios.

La medicina es una de las áreas de mayor desarrollo en innovaciones educativas en cuanto a formación profesional. Sin embargo, resta recorrer parte importante del camino que conduce hacia una verdadera profesionalización docente. En este sentido, la baja dedicación y el aumento de la carga horaria destinada a la tarea asistencial pueden atentar contra dicha profesionalización.

De todas formas, la función docente, más aún si se desempeñan a la par cargos de responsabilidad institucional, suele demandar toma de decisiones educativas. Con frecuencia, estas conllevan implementar propuestas alejadas de las tradicionales o de aquellas en desarrollo ya que, en diversas circunstancias, las demandas sociales mutan y exigen profesionales con otras orientaciones.

En tales situaciones, las innovaciones a introducir generan cambios no siempre sencillos de llevar adelante. Así como la literatura científica contribuye en la toma de decisión para la práctica clínica, en el área de educación en ciencias de la salud son varios los actores que participan en las decisiones y en su ulterior implementación.

Más allá de que quienes intervienen en la toma de decisiones pueden o no tener vasta experiencia como docentes, sólida formación pedagógica, mayor o menor nivel de involucramiento en la dinámica institucional, no son pocas las veces que estos se hallan insertos en las propias estructuras de la organización.

En ese sentido, una de las demandas más relevantes recibidas a la fecha por los formadores es la necesidad de reforzar, desde el aspecto formativo del futuro profesional, el desarrollo de la inquietud por el conocimiento y la investigación científica para que puedan encauzar críticamente su práctica y colaborar con la producción de información relevante para la posterior toma de decisiones de los gestores sanitarios. Resulta de interés señalar, según se reseñó al comienzo, que hoy más que nunca la sociedad y la práctica médica reclaman profesionales que respalden sus decisiones en datos y resultados demostrables. En la actualidad, la práctica profesional se encuentra asediada por los avances tecnológicos en diagnóstico y tratamiento, nuevos fármacos, nuevos aparatos... y la educación médica no puede estar ajena a ello.

Dado que, además, la tecnología invade y atraviesa también cuestiones relacionadas con la práctica docente y con la enseñanza: ¿cuánto es dable avanzar al respecto y de qué dependen esos avances?

Ha sido mencionado ya en este libro que determinados autores sostienen que la MBE va en detrimento de la libertad individual de cada médico. No obstante, es de suma importancia acordar que los recursos en salud, en cualquier país del mundo, son finitos y deben ser utilizados con conciencia colectiva más allá de la fuerza que toma el aspecto ético del ejercicio profesional y que supone fundamentar las decisiones en alternativas científicamente avaladas.

De no contar con estudios científicos que avalen la toma de decisiones quedaríamos expuestos, probablemente, a un desbalance entre efectividad y costo y conductas prestacionales no exentas de riesgos. Con todo, una limitación significativa a la hora de la información es que sin dudas pueden existir riesgos de sesgos en las publicaciones a la luz de que —es igual de cierto— los investigadores se hallan presionados por publicar asociaciones

y resultados positivos, buscados y apreciados por las revistas de mayor impacto, donde el investigador desea publicar.

¿Qué debemos hacer como educadores médicos? La sociedad del presente requiere profesionales con real compromiso con la carrera elegida y con la sociedad, capaces de producir cambios, buscar soluciones y proponer alternativas. Estas nuevas demandas implican nuevas estrategias de formación y ellas, a su vez, verdaderas innovaciones curriculares.

Incluir MBE en la formación profesional es esencialmente relevante para formar profesionales comprometidos con los cambios a llevar adelante. Así, también se torna imprescindible trabajar sobre la producción de firme evidencia acerca de cambios curriculares o de pequeñas innovaciones y experiencias exitosas en educación médica a fin de tomar decisiones sustentadas en otras experiencias.

La investigación educativa en educación médica constituye un área cuyo desarrollo se ha visto incrementado en los últimos años a través de reuniones internacionales y sociedades científicas dedicadas al tema. Ello obedece a que las decisiones innovadoras deben apoyarse en datos lo más objetivos posibles, producto de procesos de investigación rigurosos sistemáticamente socializados.

A pesar de ello, importante es reseñar la nada sencilla empresa de diseñar e implementar innovaciones por la comprensible, aunque injustificable, aprensión de modificar la propia práctica y de emprender cambios no suficientemente probados ni científicamente validados.

Según Tejedor (2008), «la incorporación de las experiencias realizadas y contadas, de la forma más detallada posible, por los profesores iría conformando el Banco de Evidencias».

Si bien no es una limitación menor carecer de evidencia científica en favor de los distintos procesos innovadores que diferentes instituciones educativas tienen en marcha y cuyo principal objetivo es mejorar la calidad y la pertinencia de la formación médica, es menester no amilanarse, proseguir en la tarea y propender a cultivar una educación médica basada en la evidencia.

5 El paso a paso de la MBE

ETAPAS DE LA APLICACIÓN DE LA MBE

En los capítulos previos nos limitamos a establecer el concepto de MBE y su objetivo de reducir la brecha entre ciencia y tecnología y el conocimiento práctico del profesional al permitir que las decisiones médicas se basen en la mejor evidencia científica. Para lograr este fin, el proceso implica conocer los niveles de evidencia y desarrollar cuatro etapas que favorecen alcanzar el objetivo adecuadamente.

Con el propósito de hacer más dinámica la integración de etapas de la MBE, vamos a contextualizarlas empleando una situación problema.

Caso clínico

Marta, de 35 años de edad, es oriunda de una zona rural del norte de Argentina. Asiste a la consulta preocupada porque a su hermana, de 37 años, le hicieron recientemente el diagnóstico de enfermedad de Chagas y ella quiere saber si también tiene la misma patología. Los resultados serológicos solicitados para corroborar la infección por *T. cruzi* resultan positivos, por lo cual el médico valora afección orgánica constatando únicamente mediante electrocardiograma extrasístoles ventriculares aisladas. En este contexto, la paciente consulta si existe un tratamiento para la enfermedad de Chagas y, si es así, si el mismo va a mejorar su problema cardíaco.

NIVELES DE EVIDENCIA

Cuando hacemos mención a niveles de evidencia, nos referimos al tipo de trabajo realizado para satisfacer un cuestionamiento científico. Los diseños de los estudios se clasifican con relación a 4 ejes:

- Finalidad: analítica o descriptiva. Esto supone que el objetivo del trabajo puede ser detallar una situación sin buscar asociaciones (reporte de caso, serie de casos) o plantear el contraste de hipótesis (ecológico, transversal,

caso y control, cohorte, ensayo clínico, valoración de pruebas diagnósticas, metaanálisis o revisión sistémica).

- Asignación de los factores de estudio: experimental u observacional. El primero implica la intervención por parte de los investigadores, quienes asignan el factor de estudio controlando el entorno, por ejemplo, la evaluación de la efectividad de una nueva droga (intervención) con relación a otra ya empleada o un placebo (control) para el tratamiento de una enfermedad. En el caso de los estudios observacionales, el factor de estudio no es controlado por el investigador, y se utilizan generalmente para analizar asociaciones, predictores o la «fuerza» con la que el factor influye en el desarrollo de un evento.
- Secuencia temporal: de seguimiento (longitudinal) o de una sola observación (transversal). En el caso de la investigación longitudinal, hay un control temporal de cada individuo que forma parte de la muestra del estudio. Por ejemplo, si queremos saber si un nuevo fármaco logra un mejor control de la tensión arterial en comparación con el Enalapril en un plazo de 5 años, se pautarán en el estudio los distintos tiempos en los que se controlará la tensión arterial hasta cumplir el seguimiento. En tanto, en el trabajo transversal el análisis de la muestra del estudio se efectúa una sola vez. Si queremos saber si el grosor del tejido adiposo epicárdico se relaciona con alteración de la tasa de filtrado glomerular, a cada individuo incluido en la muestra se le hace por única vez la determinación del grosor del tejido epicárdico y del filtrado glomerular; luego de esto no se vuelve a controlar al participante del estudio.
- Cronología de los hechos respecto del inicio del estudio: prospectiva (hacia adelante) y retrospectiva (hacia atrás). Las series temporales, que implican seguimiento, pueden analizarse de dos formas dependiendo de si las personas incluidas en el estudio han desarrollado o no el evento de interés. Por ejemplo, si queremos valorar si fumar es un factor de riesgo para el desarrollo de enfermedad pulmonar obstructiva crónica (EPOC), podemos incluir personas sin dicho evento (desarrollo de EPOC) con y sin el factor de estudio, es decir, fumadoras (muestra de estudio) y no fumadoras (muestra de control). Como aún no presentan el evento de interés, se realizará el seguimiento de los individuos por un plazo de tiempo, y esto se denomina estudio prospectivo.

En caso contrario de que no podamos hacer el seguimiento de la muestra, se puede efectuar un estudio retrospectivo, esto es, se incluyen personas que ya presenten EPOC y, mediante entrevista y análisis de sus historiales médicos, se analiza la fuerza de asociación con el hábito tabáquico.

Es importante remarcar que en situaciones uno puede leer: «se realizó un trabajo transversal prospectivo» o «de inclusión prospectiva», y esto refiere a cómo es la inclusión de las personas que conforman la muestra y

no a la relación temporal entre factor y evento. Recordemos que el trabajo transversal es el análisis de un momento.

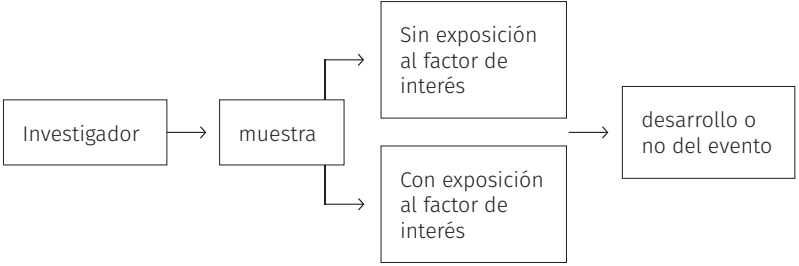


FIGURA 2. SERIE TEMPORAL PROSPECTIVA (COHORTE)

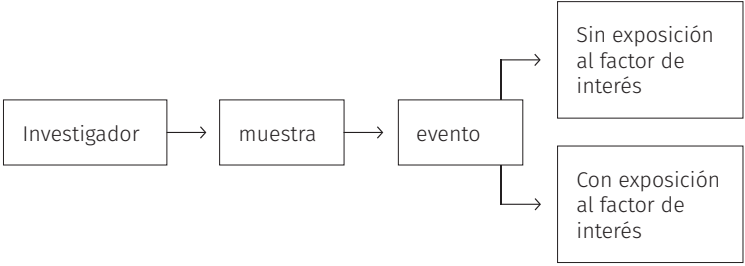


FIGURA 3. SERIE TEMPORAL RETROSPECTIVA (CASO Y CONTROL)

En el cumplimiento de estas características se constituye la pirámide del nivel de evidencia sobre el cual se enmarca la respuesta a la pregunta clínica.

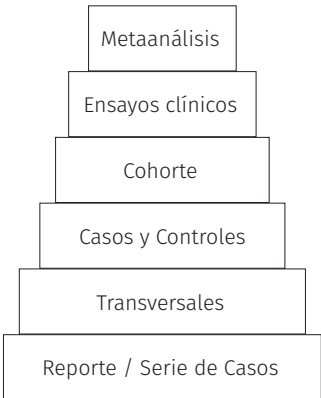


FIGURA 4. PIRÁMIDE DEL NIVEL DE EVIDENCIA

Si bien en los próximos capítulos detallaremos los tipos de estudio mencionados, se puede resumir la utilidad de cada trabajo científico respecto de la pregunta clínica de la siguiente forma:

TABLA 2. UTILIDAD DE LOS TRABAJOS CIENTÍFICOS

Marco general	Ejemplo	Tipo de estudio
¿Hay un caso o casos similares a mi paciente?	¿Estará descrita la reactivación de la enfermedad de Chagas posquimioterapia?	Reporte de Caso o Serie de Casos (descriptivo)
¿Existe asociación entre dos o más variables?	¿Se asociará el grosor del tejido adiposo epicárdico con la tasa de filtrado glomerular?	Estudios Transversales (de asociación o correlación)
¿Será que el evento que tiene mi paciente se debe a que presenta determinado antecedente? (chance)	¿Será que mi paciente sufre de enfermedad pulmonar obstructiva crónica por el hecho de ser tabaquista?	Estudio Casos y Controles (retrospectivos de asociación o correlación, permiten determinar chance)
¿Será que el antecedente que refiere mi paciente influyó en el desarrollo de su evento? (factor de riesgo)	¿Será que hipertensión arterial es un factor de riesgo para síndrome coronario agudo?	Estudio Cohorte (prospectivos de asociación o correlación, permiten determinar riesgo)
¿Es útil el tratamiento para la enfermedad?	En mi paciente con accidente isquémico transitorio, ¿será conveniente tratarlo con estatinas a fin de evitar nuevos episodios?	Ensayos clínicos (prospectivos, permiten determinar la efectividad de una intervención en relación con un evento de interés)
¿Es útil determinado método diagnóstico?		Estudio transversal o de cohorte con <i>gold standard</i>

Contextualizados los niveles de evidencia, podemos continuar con las etapas del proceso de la MBE en la toma de decisiones.

ETAPAS DE APLICACIÓN

Poner en práctica la MBE implica llevar a cabo una serie de acciones que favorecen su adecuado desarrollo. En primer lugar, se debe definir adecuadamente el problema, transformarlo en un interrogante clínico contestable y proceder

luego a la búsqueda en las bases de información pertinentes al nivel de la pregunta (básica o avanzada). Una vez resuelta la búsqueda, es necesario realizar un análisis crítico de los textos científicos encontrados (aplicando conocimientos bioestadísticos y de metodología de la investigación clínica) para dar respuesta a nuestra pregunta. Si la duda persiste, debe reiniciarse el proceso generando nuevamente la pregunta clínica.

Paso 1: Hacer de un dilema médico una pregunta que pueda tener respuesta

Desde que iniciamos nuestra etapa como estudiantes de la carrera de Medicina, nos planteamos diversos interrogantes que motivan el desarrollo y construcción del conocimiento. La necesidad de saber nos impulsó mediante el desarrollo de preguntas clínicas a realizar búsquedas bibliográficas y/o consultas a expertos para tratar de darnos respuestas. Como nadie posee un saber absoluto (como tampoco verdades absolutas), este ciclo se da en forma continua y repetitiva también a lo largo del ejercicio médico (solo que en vez de llamarse «estudiante de medicina», una vez graduado pasa a ser MBE).

De igual forma, ante la necesidad del saber, se genera la primera etapa de la MBE, la formulación adecuada de la pregunta clínica a partir de la situación problema del paciente. El planteo del interrogante debe cumplir con determinadas características. En primer lugar, tiene que estructurarse en forma contestable, es decir, que sea factible de tener respuesta, lo cual implica que su formulación debe contextualizarse desde el conocimiento previo sobre el tema a buscar. Esto disminuye las posibilidades de encontrar respuestas equivocadas o imprecisas.

Se pueden establecer dos tipos de preguntas clínicas:

- Básicas o generales (*background*): surgen de aspectos básicos sobre una determinada condición clínica. Se estructuran con dos componentes esenciales, raíz «quién», «cuál», «dónde», «cómo» o «por qué» y el verbo. Considerando el caso clínico citado, un ejemplo de este tipo de pregunta sería: ¿cuál (raíz) es (verbo) el agente etiológico de la enfermedad de Chagas? Como se puede apreciar, esta clase de interrogante es de un aspecto básico cuya respuesta puede obtenerse en textos de referencia como libros o manuales médicos.
- Avanzadas, específicas o de actualización (*foreground*): en este caso, las preguntas son orientadas a aspectos puntuales de situaciones clínicas, principalmente tratamiento, diagnóstico, pronóstico y prevención. La estructura de este tipo de preguntas se compone de paciente o problema clínico, intervención de interés (pudiendo referirse a validez de pruebas

diagnósticas, tratamientos, utilidad de métodos complementarios, etc.), Comparación de intervenciones (cuando sea pertinente) y el *outcome* (resultado). Esta modalidad de búsqueda conforma lo que se denomina estrategia PICO.

Si consideramos el caso de ejemplo, la pregunta clínica podría ser: ¿en personas con miocardiopatía chagásica (problema), el tratamiento con benznidazol (intervención) es efectivo (*outcome*)?

En este caso, la respuesta al interrogante se va a encontrar en artículos de revistas científicas.

Es importante remarcar que todas las situaciones problemas pueden inducir a la conformación de preguntas clínicas (sean básicas o específicas) en las diversas áreas de las ciencias médicas.

Paso 2: Realizar la búsqueda bibliográfica

Luego de establecer la pregunta clínica se procede a realizar la búsqueda bibliográfica. Debe existir coherencia entre la pregunta y el método mediante el cual se realiza este paso, de lo contrario tal vez no encontremos la mejor respuesta a nuestro interrogante.

Existen dos bases de publicaciones científicas biomédicas principales, *Medline* y *Embase*. La primera pertenece a la Biblioteca Nacional Médica de Estados Unidos (*National Library of Medicine*), que indexa a más de 5500 revistas médicas desde 1966 y cuenta en la actualidad con más de 10 millones de referencias. Por su parte, *Embase* (*Excerpta Medica Database*), si bien cuenta con un mayor número de *journals* indexados (aproximadamente 8400), el número de referencias es un poco menor, estimado en 8 millones desde 1974.

El método de búsqueda ideal es mediante ambas bases de datos. Esto se debe a que existe solamente un 34 % de resultados en común; sin embargo, para acceder a *Embase* se requiere de una suscripción, mientras que la búsqueda en *Medline* es sin costo a través de PubMed.

Las direcciones de Internet para ingresar a dichas bases son:

Medline: www.pubmed.gov

Embase: <http://www.elsevier.com/solutions/embase-biomedical-research>

Además de los sitios mencionados, se pueden emplear otros buscadores con acceso a la base de datos *Medline* y a artículos no indexados en la misma, por ejemplo *Tripdatabase* (www.tripdatabase.com). Independientemente de cuál se emplee, es importante destacar que siempre la búsqueda debe ser planteada de manera adecuada, en caso contrario la respuesta a nuestra pregunta clínica puede llegar a ser incorrecta. A fin de facilitar este proceso, existe lo que se denomina «vocabulario MeSH», definido como el grupo de

términos empleados por la *National Library of Medicine (Medline)* que deben ser empleados por quienes escriben los artículos científicos como «palabras clave» de búsqueda. Esta base de términos, MeSH database (*Medical Subject Headings*), es de libre acceso en línea.

En forma esquematizada, los pasos para realizar la búsqueda de un artículo se pueden resumir en:

- 1) Planteo de pregunta clínica: ¿en personas con enfermedad miocardiopatía chagásica (problema/paciente), el tratamiento con benznidazole (intervención) es efectivo (*outcome* —resultado—)?

Nos encontramos con una pregunta de tipo avanzado. Considerando que la misma es sobre la eficacia de una intervención terapéutica, el artículo científico ideal a buscar es un ensayo clínico (o, subiendo el nivel de evidencia, metaanálisis/revisión sistémica).

- 2) Ingreso a la base de datos, por ejemplo, *Medline*.

Si bien existen diversas metodologías de búsqueda en el sitio, es recomendable para los que se inician usar la modalidad avanzada de búsqueda. Mediante la misma se puede seleccionar si se van a usar uno o más términos (palabras clave) y en qué parte del cuerpo del artículo científico se deben encontrar. También se puede buscar por autores, año, tipo de trabajo científico, si es en humanos o en modelos animales de experimentación, entre otras cosas.

Con relación a la pregunta clínica pautada en el punto 1, mediante la opción avanzada de búsqueda de PubMed podemos establecer la siguiente metodología:

- Title/Abstract: Chagas' Cardiomyopathy
- Title: Benznidazole
- All Fields: Clinical Trials

De esta forma, le estamos diciendo que nos busque en su base de datos todos los artículos que contengan los términos cardiomiopatía chagásica (en el título o resumen), benznidazole (en el título) y ensayo clínico (en cualquier parte del texto).

También podría recurrirse a herramientas diseñadas para aplicar la estrategia PICO directamente en la búsqueda, como la opción «PICO search» de www.tripdatabase.com o su equivalente para PubMed (<http://pubmedhh.nlm.nih.gov/nlmd/pico/piconew.php>).

- 3) Valorar si los resultados son coherentes a nuestra pregunta clínica o reiniciar el proceso. En cuanto a nuestro planteo, el resultado obtenido satisface los criterios de búsqueda y el artículo puede ser de utilidad para responder a nuestro interrogante clínico.

- Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy, publicado en *New England Journal of Medicine*, por los autores: Morillo, C.A.; Marin-Neto, J.A.; Avezum, A; Sosa-Estani, S.; Rassi, A. Jr.; Rosas, F.; Villena, E.; Quiroz, R.; Bonilla, R.; Britto, C.; Guhl, F.; Velazquez, E.; Bonilla,

L.; Meeks, B.; Rao–Melacini, P.; Pogue, J.; Mattos, A.; Lazdins, J.; Rassi, A; Connolly, S.J.; Yusuf, S.; BENEFIT Investigators.

Como dijimos, existen diversas modalidades para realizar el proceso de búsqueda. En este apartado hemos tratado de resumir en forma esquemática el desarrollo del mismo. Sugerimos la lectura de los siguientes artículos que le van a permitir interiorizarse en detalle de cómo realizar una adecuada búsqueda científica.

- Melnyk, B.M.; Fineout–Overholt, E. (2005). Making the case for evidence–based practice. In Melnyk, B.M., Fineout–Overholt, E. (eds.). *Evidence–based practice in nursing and healthcare: a guide to best practice* (pp. 3–24). 1st. ed. Philadelphia: Lippincott Williams and Wilkins.
- Da Costa Santos, C.M.; De Mattos Pimenta, C.A.; Nobre, M.R. (2007). The PICO strategy for the research question construction and evidence search. *Rev. Lat. Am. Enfermagem*, 15(3 May–Jun), 508–511.
- Tutorial PubMed: <https://www.nlm.nih.gov/bsd/disted/pubmed.html>

En conclusión, una vez encontrados los artículos científicos propios a nuestra situación problema se debe determinar su validez y utilidad. Es recomendado para esto, contar con conocimientos básicos de bioestadística que permitan comprender adecuadamente los datos discutidos en el artículo científico (lectura crítica).

Paso 3: Evaluar críticamente los resultados de la búsqueda

Como fue mencionado previamente, saber algunos conceptos de bioestadística (ver Capítulo 6) y cuál es el nivel de evidencia requerido para satisfacer adecuadamente a nuestra pregunta clínica son aspectos útiles para evaluar críticamente los artículos. Pero, ¿a qué nos referimos con «evaluar críticamente» el texto científico?

Esta etapa implica, en primer lugar, conocer en qué revista científica se ha publicado el trabajo y su nivel respecto de otras del mismo tema. En segunda instancia, desglosar el artículo y analizar su coherencia interna (validez interna). Para el primer paso, existen diversos indicadores que tratan de medir la calidad de la revista (*journal*). Entre ellos, los dos más empleados son:

- Factor de impacto: es el cociente entre el número de citas que han recibido en ese año los documentos publicados en los 2 años anteriores y el número de documentos publicados por la revista en esos 2 años. Es un indicador que permite comparar revistas y evaluar la importancia relativa de cada una dentro de un mismo campo científico, donde el investigador comunica sus resultados. Se considera que la revista científica en el campo de las ciencias médicas es de relevancia cuando su Factor de Impacto es ≥ 2 .

- *SCImago Journal Rank*: igual que el anterior, valora la calidad y reputación de la revista científica y determina la posición de una revista en relación con todas las de su área. Se divide en 4 partes iguales un listado de revistas ordenadas de mayor a menor índice *SJR* (que incluye al factor de impacto) y cada una de estas partes será un cuartil. Se considera que una revista científica es importante dentro del área de las ciencias médicas si se encuentra en el cuartil 1 ó 2.

Los factores de impacto de las revistas se pueden buscar en el sitio propio de los *journals* o en buscadores específicos como <http://www.bioxbio.com/if/>. Por su parte, en el sitio de Internet <http://www.scimagojr.com/index.php>, se puede encontrar el *SCImago Journal Rank*. Si bien el artículo de interés puede estar publicado en una revista prestigiosa, no quiere decir con seguridad plena que el mismo no presente puntos cuestionables en cuanto a su validez.

Esto nos lleva a la segunda etapa, analizar que exista una relación adecuada entre la hipótesis del trabajo, sus objetivos, la metodología aplicada, y que los resultados reflejen lo que se pretendió estudiar. Otro factor importante a evaluar es la presencia de sesgos. Vamos a detenernos en este último punto. En el proceso de investigación pueden existir errores aleatorios que son inherentes a la variabilidad de la muestra con la cual se va a desarrollar el trabajo científico. Por el contrario, los sesgos son propios de la metodología empleada. Existen varios tipos de ellos, que comprenden básicamente 4 grandes grupos:

- **Sesgo de selección**: sucede cuando los procesos empleados para conformar la muestra o para realizar el seguimiento no concuerdan con el tipo de estudio y los objetivos planteados.
- **Sesgo de medición**: las herramientas para realizar la recolección de la información observada o la forma de medir la misma no son adecuadas. Pueden generar, por ejemplo, clasificación incorrecta de los sujetos o disparidad de datos intra e intergrupos. Un sesgo dentro de este punto es el sesgo de memoria, que ocurre en los trabajos retrospectivos, y refiere al olvido por parte de los sujetos de hechos inadvertidos por ellos pero que pueden ser muy relevantes en el impacto del trabajo.
- **Sesgo de confusión**: acontece cuando existe una o más variables que modifican el impacto del factor de riesgo con relación al desarrollo del evento, y no son determinadas.
- **Sesgo de publicación**: es muy frecuente en las revisiones sistemáticas y sucede cuando se considera que los estudios publicados sobre un tema son realmente todos los trabajos científicos realizados. Retomaremos este tema más adelante.

Por más rigurosos que sean los investigadores que llevan a cabo el estudio, hay que considerar que es imposible evitar absolutamente todos los sesgos.

Saber que existen, en un sentido amplio, permite al lector tratar de identificarlos para evaluar si su presencia influye en los resultados y conclusión del trabajo científico.

Vale recordar que, al estar publicado el *paper*, el mismo ha tenido que sortear los obstáculos propios de la editorial de la revista, lo que incluye, en la gran mayoría de los casos, la revisión por parte de pares (otros investigadores que evalúan la calidad del artículo). Por otra parte, si bien nos da «seguridad» que el artículo se encuentre publicado en revistas de alta relevancia científica, no debemos excluir automáticamente trabajos de otros *journals*. Siempre corresponde ver si cumple con la validez interna para responder a nuestra pregunta clínica. Si esto es así, nos queda solo un paso para poder concluir el proceso de MBE.

Paso 4: Aplicar la información obtenida a nuestra situación problema

Ya estamos en la última etapa de la aplicación de la MBE. Como los anteriores, este paso es crucial. De concretarse satisfactoriamente, hemos resuelto nuestra pregunta clínica (con el agregado de que en ocasiones promueve el desarrollo de nuevos planteos y búsquedas). Por el contrario, si el texto científico no es aplicable a nuestra situación problema o no logra responder a nuestra pregunta, hay que volver a empezar y rever si no nos equivocamos al formular la pregunta o al analizar críticamente el trabajo a fin de subsanar nuestro error.

Cuando decimos «aplicar la información obtenida a nuestra situación problema» no hablamos de trasladar textualmente la conclusión del texto científico a nuestro paciente sino que, sobre la base de la experiencia y en un contexto dialógico con el paciente, debemos establecer la relación riesgo/beneficio del accionar médico adquirido durante el proceso de MBE. Este proceso se denomina «validez externa». Acá ya no nos interesa tanto el valor de p , que nos dice si es significativo el análisis estadístico. Lo que nos importa es la significancia clínica, es decir, la relevancia que tiene el efecto observado por los investigadores para nuestra situación problema en particular.

¿Cómo? ¿No todos los hallazgos significativos son importantes? Para clarificar este punto, veamos el siguiente ejemplo. Supongamos que nos surge la duda acerca de si un nuevo antihipertensivo (droga A) es mejor que el Enalapril para el control de la tensión arterial en personas mayores de 70 años con antecedente de infarto agudo de miocardio, siendo que la droga A cuesta 4 veces más. Procedemos a realizar la búsqueda pertinente y encontramos un ensayo clínico publicado en una prestigiosa revista con una validez interna óptima. La conclusión del ensayo clínico nos dice que la nueva droga controla mejor la tensión arterial, reduce 3 mmHg más que el Enalapril la tensión arterial sistólica (el test estadístico es significativo, $p < 0.000000001$ —y tantos

ceros como guste delante del 1—). Las preguntas que deben surgir luego de la lectura son: ¿vale la pena realmente el nuevo fármaco? ¿Los 3 mmHg de diferencia con el Enalapril en el control de la tensión arterial sistólica hacen necesario que le reemplace el tratamiento a mi paciente?

Estos interrogantes son parte de la reflexión luego de leer el artículo. Siempre debemos reflexionar sobre la validación externa del texto científico. Para cada tipo de trabajo hay preguntas orientativas específicas respecto de la validez externa que trataremos en los capítulos correspondientes; ahora vamos a resumir algunos puntos que guían la lectura del *paper*.

- 1) ¿Quiénes son los autores del artículo? ¿Presentan algún conflicto de interés?
- 2) De presentar conflicto de interés los autores, ¿el mismo se relaciona con el objeto de estudio del trabajo?
- 3) ¿Qué tipo de trabajo es, a qué nivel de evidencia corresponde?
- 4) ¿Los objetivos del estudio contextualizan nuestra pregunta clínica?
- 5) ¿Es analítico u observacional?
- 6) ¿Dónde se realizó el trabajo?
- 7) ¿Se evitaron errores metodológicos del desarrollo del estudio (sesgos)?
- 8) ¿Los pacientes incluidos en este trabajo son parecidos a mi paciente?
- 9) Las conclusiones, ¿tienen significancia clínica? ¿Son aplicables a mi situación problema?

La resolución de estas preguntas va a posicionarnos en el marco adecuado para que, sobre la base de nuestro saber, analicemos la eficacia y el beneficio que obtendríamos al aplicar la respuesta a nuestra pregunta clínica en el caso problema disparador, siempre en una postura dialógica con el paciente.

6 Nociones básicas de bioestadística

Antes de dar inicio con la descripción de los diversos niveles de evidencia, es necesario desarrollar conceptos generales de bioestadística para lograr una mejor comprensión de los textos científicos. No explicaremos todas las pruebas estadísticas que existen sino que nos limitaremos a las empleadas habitualmente.

CONCEPTOS GENERALES

Cuando los investigadores se plantean cuestiones de interés a ser analizadas están conformando su hipótesis de trabajo. La misma se debe enfocar desde una perspectiva nula en tanto defina que no existe relación entre el objeto de estudio (variable dependiente) y aquellos factores de la realidad que pueden influir en él (variables independientes). A partir de la hipótesis se estructuran los objetivos del trabajo que representan, las acciones que se llevarán a cabo para contrastar el supuesto teórico. En concreto, se puede decir que la hipótesis de estudio es el análisis de la relación entre factor y evento de interés.

Se preguntará la razón de que se deba plantear la hipótesis nula (H_0). Pues bien, como se verá más adelante, los análisis estadísticos son utilizados para contrastar H_0 aceptando o rechazando a la misma. Es decir, el valor estadístico permite considerar verdadera a la H_0 o alejarnos de ella hacia la hipótesis alternativa (H_1).

Sobre la base de la hipótesis y objetivos se pauta el tipo de trabajo a realizar. Como referimos previamente, los niveles de evidencia científica corresponden al tipo de estudio que da respuesta a la pregunta clínica. Es decir, el diseño del trabajo científico mediante el cual se evalúa la realidad que se investiga. Cabe destacar que existen estudios observacionales descriptivos que no requieren de una hipótesis debido a que el objetivo de estos trabajos es simplemente identificar características del objeto de estudio.

TABLA 3. TIPO DE ESTUDIO

Tipo de Estudio	
Experimental	No experimental
Ensayo clínico	Estudio ecológico Estudio de prevalencia Estudio caso y control Estudio cohorte

Por otra parte, además de establecer el tipo de trabajo científico, la hipótesis y los objetivos contextualizan las variables que se emplean en el trabajo. Las variables se definen como el valor de un conjunto determinado y pueden ser:

- Cuantitativas o numéricas: son las que nos permiten hacer operaciones aritméticas. Es posible agruparlas en dos categorías: 1) continuas, cuando admiten modalidad intermedia entre sus valores, por ejemplo, peso, ya que entre 30 y 40 kilos existen valores mesurables; 2) discretas, al no admitir modalidad intermedia entre sus valores, y un ejemplo de este tipo de variable es «número de hijos», debido a que entre 3 y 4 hijos no existe un valor mesurable (no se puede tener 3 hijos y medio).
- Cualitativas: son nominales u ordinales. En el primer caso, la variable adopta un valor sin orden jerárquico, mientras que en el segundo es necesario el orden. Por ejemplo, el grupo sanguíneo es A, B, AB o O (sin orden entre ellos), mientras que la categorización del Índice de Masa Corporal en normal, sobrepeso y obesidad refiere una escala de orden entre los valores (0 normal, 1 sobrepeso, 2 obesidad).

Contextualizado el primer paso del desarrollo de un trabajo científico en el planteo de la hipótesis, objetivos, diseño de trabajo y variables a incluir, es común leer en los *papers* de investigación médica realizados en personas que se selecciona una parte de la población, es decir, se conforma una muestra (subconjunto). Esto sucede por la imposibilidad que implica analizar el conjunto por completo. Imaginemos que nos planteamos saber si las personas mayores de 45 años presentan riesgo de desarrollo de diabetes como consecuencia de la toma diaria de estatinas. Solo en Estados Unidos, más de 51 millones de individuos durante 2009 se encontraban bajo tratamiento con estos fármacos, por lo cual estudiar nuestra hipótesis en la población completa no es factible.

Es entonces cuando entra en acción la estadística. La misma se define como los procedimientos empleados para concretar la recolección y análisis de datos. Una de sus formas es la descriptiva, que es cuando se utilizan datos

numéricos o gráficos que resumen la representación de un grupo de datos, por ejemplo, «la edad fue de $35,3 \pm 1,3$ años y el 54,7 % fue mujer». En esta situación únicamente se detallan las variables, siendo el promedio de edad (\pm el desvío estándar) y la proporción de mujeres. Como se puede observar, este tipo de estadística no tiene como fin contrastar una hipótesis porque solamente es el examen descriptivo de variables (e incluso no es necesario la existencia de una hipótesis para realizar estadística descriptiva).

Ahora, ¿qué pasa si queremos valorar si nuestra variable dependiente presenta relación con variables independientes? Es entonces cuando se hace necesario realizar estimaciones, cálculo de probabilidades, riesgos, chances, o predicciones. Esto se denomina «estadística inferencial». En el ejemplo previo no existe una hipótesis, pero para hacer uso de la estadística inferencial es indispensable ya que los análisis estadísticos nos permiten aceptar o no la hipótesis nula.

Un concepto fundamental que se debe incorporar en la práctica de la estadística inferencial es la significación estadística, también conocida como el valor de p . Se puede definir como la probabilidad de que lo observado sea real y no consecuencia de chance. Por ejemplo, si leemos en un trabajo «el peso corporal se correlacionó en forma moderada con la edad ($r=0,79$, $p<0,03$)», se está planteando lo siguiente:

- 1) H_0 : no existe relación entre edad y peso corporal.
- 2) El test estadístico empleado (correlación de Pearson $-r-$) rechaza la H_0 , infiriendo que las variables edad y peso se relacionan positivamente (a mayor edad se observa mayor peso), con una significancia estadística (p) $<$ a $0,03$ (escasas probabilidades de que nuestro resultado estadístico se deba al azar). Esto representa a la hipótesis alternativa.

Otro término que suele acompañar los resultados estadísticos, además de p , es el Intervalo de Confianza (IC), definido como el rango que contiene al valor observado. Suele emplearse el IC de 95 %, lo que implica que se analiza cuál es el margen necesario en el que se encuentra con un 95 % de probabilidad el resultado del test estadístico.

En resumen, al plantearse un supuesto teórico sobre un aspecto de la realidad que se desea investigar se construye una hipótesis que enmarca a los objetivos (actividades concretas) que se van a llevar a cabo. Para cumplir con estos se establecen las variables de estudio y los mecanismos de recolección y análisis de datos. De esta manera se conforma la coherencia interna de un trabajo científico.

Hipótesis \longrightarrow Objetivos \longrightarrow Diseño \longrightarrow Muestra (n) \longrightarrow Análisis \longrightarrow Resultados

LAS PRUEBAS ESTADÍSTICAS

La mayoría de los trabajos científicos médicos se realiza mediante la selección de una muestra de la población debido a las limitaciones que implica analizar al total. Sin embargo, como el objetivo es estudiar un fenómeno que sucede en el conjunto, lo ideal es que los aspectos de nuestra muestra (variables) sean lo más parecidos a la de la población.

Imaginemos que vamos a realizar un estudio incluyendo estudiantes universitarios de una universidad particular, es decir, miles de personas, con el propósito de valorar si la altura varía con relación a las distintas carreras. La H_0 sería en este caso que no existe variación de la altura de los estudiantes respecto de la carrera universitaria en la cual están inscriptos. Como todos los estudiantes tendrán una altura determinada, habrá una frecuencia para cada altura. Es decir que en la población de estudiantes universitarios se repartirá entre las distintas alturas; en otras palabras, la altura tendrá una distribución de probabilidad determinada para la población en estudio. Ante la limitación para estudiar a todos los alumnos de la universidad, se selecciona una muestra, un subgrupo en el cual vamos a medir la altura y así intentar conocer la distribución de la población. Si la muestra es suficientemente significativa, podremos llegar a una aproximación de la distribución de la población.

¿Cómo se puede saber si la muestra representa a la población en relación al efecto que se va a explorar? Hay cálculos matemáticos que permiten estimar el tamaño muestral (n) que se requiere para poder valorar la hipótesis del trabajo. No es de interés de este libro explicar cada una de las fórmulas, pero sí dos conceptos indispensables que nos permiten evaluar si el n es adecuado para contrastar la hipótesis planteada y extrapolar los resultados del trabajo científico. Nos referimos a los errores tipo I (α) y tipo II (β).

- Error tipo I (α): es cuando se rechaza la hipótesis nula siendo que en realidad esta es cierta. Es decir, aceptamos que existe relación entre nuestro objeto de estudio y las variables independientes cuando en realidad no es así (falso positivo). Se denomina a este error como «significado estadístico (p)». Se establece el error $\alpha = 0,05$, por lo que la máxima probabilidad de rechazar incorrectamente la hipótesis nula es del 5 %.
- Error tipo II (β): es cuando aceptamos la hipótesis nula siendo que en realidad esta es falsa (falso negativo). También se denomina a este error como «poder del estudio», el cual se calcula $1 - \beta$. Si el error β se establece en 0,10 estamos aceptando que trabajaremos con una chance del 10 % de no encontrar la diferencia, y el poder del estudio es de 0,90 (90 %).

Además de los errores referidos, existen otros parámetros que se emplean en el cálculo del tamaño muestral que dependen particularmente del objetivo

del estudio; por ejemplo, si es de observar una diferencia de medias, es necesario conocer la magnitud de la diferencia a detectar.

Una vez determinado el tamaño muestral, se analiza la distribución de las variables. Independientemente de cual fuera el método de análisis de distribución (prueba de Kolmogorov–Smirnov, Shapiro Wilks, contraste X^2 de Pearson, entre otras), nos interesa su conclusión. Si la probabilidad de frecuencias tiene un comportamiento normal se considera que se asemeja a la realidad y permite el empleo de análisis estadísticos denominados paramétricos, de no hacerlo solo se pueden usar test no paramétricos. A grandes rasgos, se diferencian entre ellos en que el primer tipo de análisis nos permite la extrapolación de los resultados y el análisis de parámetros de interés. Además de la distribución de la variable, es factible emplear test paramétricos cuando el tamaño muestral es superior a 30 porque se considera que la variable se distribuye con normalidad si es adecuado el número de muestra incluida (teorema central de límites) o cuando por transformaciones matemáticas se logra la normalidad.

Vamos a detenernos ahora en el desarrollo de conceptos que surgen de la distribución de variable cuantitativa continua, también denominada «distribución de Gauss». Dijimos que las variables continuas nos permiten realizar operaciones aritméticas y admiten modalidad intermedia entre sus valores.

Dentro de los parámetros de distribución que se obtienen interesan:

- Media (μ): es el valor promedio de la variable en cuestión.
- Mediana: representa el valor de posición central. Es decir, se ordena de menor a mayor la variable en estudio y el valor que se posiciona dejando 50 % por detrás y por delante de él es la mediana.
- Moda: es el valor (o los valores) que presenta mayor frecuencia en la distribución de la variable.
- Varianza: es el promedio de las diferencias de cada valor para con la media total de una variable, elevada al cuadrado.
- Desvío Estándar (DS): es la raíz cuadrada de la varianza, y es una forma de analizar cuánto se separan los datos de la media de la muestra. Mientras mayor sea la dispersión de los datos, mayor será el DS.
- Percentil: medida de posicionamiento en porcentaje de una distribución que determina valores que son menores que o igual a la cifra.
- Error Estándar de la Media (SEM): es la medida del error que se comete cuando se cuantifica el valor real de nuestra medida en la población a partir del observado en la muestra. Esto es, representa el grado de dispersión que tendría la media de la muestra si se siguen realizando muestreos. Se determina mediante el índice de DS con la raíz cuadrada del tamaño muestral $DS/(\sqrt{n})$.

En el caso de que la variable no presente distribución normal, los valores que nos interesan son mediana, y rango intercuartil (es decir, percentilos 25 y 75).

A continuación se presenta una gráfica de distribución normal en la cual se detallan la μ y los DS. En el rango de ± 1 DS se encuentra el 68 % de las observaciones, a ± 2 DS el 95 % y a ± 3 DS el 99,7 %. Considerando la simetría de la curva, la mediana y moda se corresponden con la μ .

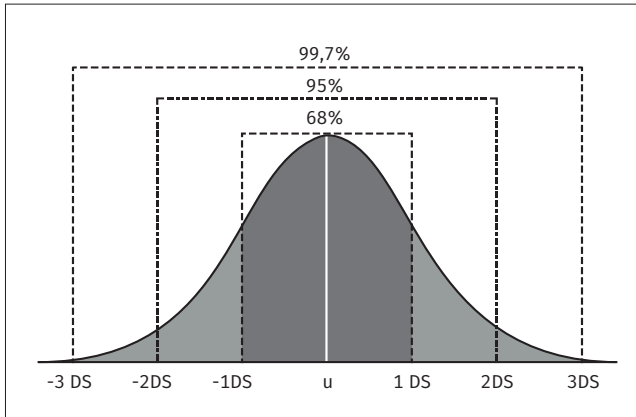


FIGURA 5. PARÁMETROS DE DISTRIBUCIÓN: DISTRIBUCIÓN NORMAL

Analicemos otra gráfica sobre la distribución de la edad en una muestra de 177 sujetos. La misma es un histograma en el que se describe la frecuencia de individuos (eje Y) con relación a la edad (eje X) acompañada de la curva de distribución que a simple vista impresiona normal.

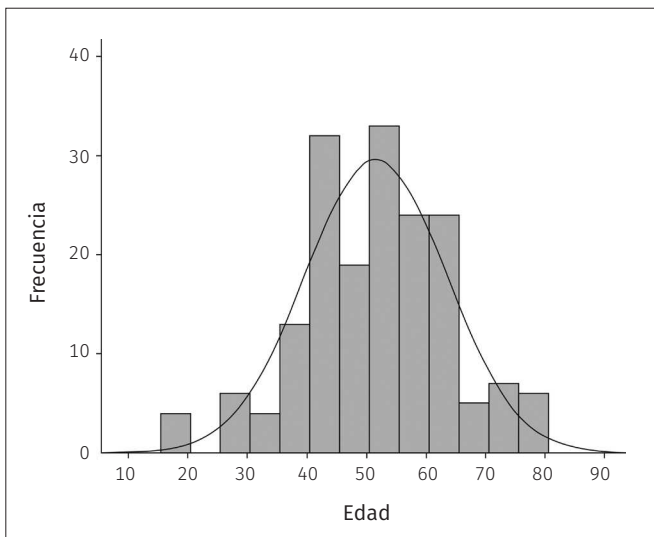


FIGURA 6. EJEMPLO DE HISTOGRAMA DE EDAD

La μ es de 51,62 y 1 DS representa 11,9 años. Por su parte, la mediana es de 52 y la moda es de 54. Si determinamos los percentiles 25, 50 y 75, los valores son 44, 52 y 60 años respectivamente. Como se puede observar, el valor μ es muy similar a la mediana, orientándonos a considerar que la distribución es normal. Su análisis mediante el test estadístico correspondiente arroja que la misma corrobora nuestra apreciación.

Ahora bien, ¿por qué no reportamos el valor de SEM? Esto es debido a que, al describir el SD, le estamos brindando a usted una idea de la dispersión de la variable entre los sujetos respecto de la media. El 68 % de la muestra tiene 51,62±11,9 años, es decir, desde 39,7 a 63,5 años. Si hubiésemos escrito 51,62±0,89 (ES) solo le estaríamos proporcionando una idea de la precisión de la edad en la muestra con relación a la población.

Rara vez se expresa SEM en las publicaciones; sin embargo, si se describe $\mu \pm SEM$, le sugerimos que transforme este último a SD para poder valorar realmente las diferencias entre sujetos o grupos.

ASOCIACIÓN ENTRE DOS VARIABLES

En conocimiento del concepto de distribución y los valores que se pueden determinar, podemos dar inicio al desarrollo de algunos análisis estadísticos que se emplean habitualmente en los textos médicos.

Existen pruebas estadísticas que son utilizadas en el análisis de lo que se denomina «muestras no relacionadas». Esto quiere decir que los datos de las variables que se estudian fueron recolectados en un solo tiempo. Por ejemplo, si se quiere valorar si existen diferencias de media entre glucemia en ayunas entre hombres y mujeres, se usa el test t-Test, donde la variable glucemia es cuantitativa continua y la variable sexo es cualitativa nominal.

TABLA 4. ESQUEMA DE ANÁLISIS DE ASOCIACIÓN

	Cuantitativas	Cualitativas
Cuantitativas	Pearson (r) – R2 Rho de Spearman	2 grupos: t-Test U M. Whitney (NP)
Cualitativas	3 o más grupos: ANOVA Kruskal Wallis (NP)	X2 (Chi-cuadrado) Exacta de Fisher (NP)

NP: no paramétricas.

Ahora, si se busca analizar si hay diferencia de media de glucemia en los distintos tiempos de una prueba de tolerancia oral de glucosa, cada paciente es su propio control en cuanto a cada determinación del valor de glucemia, es decir, glucemia en ayunas, glucemia a los 60 y a los 120 minutos. Esto se denomina «muestras relacionadas».

TABLA 5. ESQUEMA DE ANÁLISIS DE RELACIÓN

	2 mediciones	3 o más mediciones
Cuantitativas	t-Test datos apareados Prueba de Wilcoxon (NP)	Análisis de la Varianza Análisis de la Varianza de Friedman (NP)
Cualitativas	Prueba de McNemar	Q de Cochran

NP: no paramétricas.

Respecto de la valoración de asociación entre dos variables cualitativas, se puede emplear el test X^2 (Chi-cuadrado). Este proporciona un estadístico que permite contrastar la hipótesis de que los dos criterios de clasificación utilizados (las variables categóricas) son independientes. Para ello compara la frecuencia observada (es decir, las frecuencias que se han recolectado) con las frecuencias esperadas (un valor teórico que deberíamos haber hallado si los criterios de clasificación fueran independientes). Si los datos son compatibles con la hipótesis de independencia, la probabilidad asociada al X^2 será alta ($p > 0,05$). Si esa probabilidad es muy pequeña ($p < 0,05$), concluimos que las variables estudiadas están relacionadas. En el caso de que el 20 % o más de los valores esperados sea ≤ 5 , en vez de emplearse el test X^2 se utiliza el Exacto de Fisher. Hoy en día la gran mayoría de los programas de estadística arroja los dos test a fin de que el operador pueda observar cuál de ellos corresponde comunicar. Es importante recordar que la asociación no implica causa, solamente estamos observando que dos variables no presentan independencia entre ellas.

Y tal como se puede analizar la asociación entre variables cualitativas es factible explorar la correlación entre dos variables cuantitativas. En el caso de que las variables en cuestión tengan distribución normal, el test que se emplea se denomina Correlación de Pearson, que suele representarse con r . Este análisis estadístico determina asociación pero no causa. Su valor va de 0 a 1, siendo 0 la inexistencia de relación y 1 asociación perfecta. A la vez puede ser (+) cuando es directa o (-) cuando es inversa. Por ejemplo, si leemos en un trabajo «el peso corporal se correlacionó en forma moderada con la edad ($r=0,79, p<0,03$)» quiere decir que, a medida que el valor de edad se incrementa, en forma directa y moderada aumenta el peso corporal. Si, por el contrario, en la situación ejemplo, en vez de decir $r=0,79$ diría $r=-0,79$

(signo negativo antes del valor), indica que, a medida que es mayor el valor de edad, el peso corporal disminuye. En caso de que las variables no presenten distribución normal, el test estadístico que se emplea es Rho de Spearman y se lee de igual forma que la correlación de Pearson.

Con el valor de r se puede terminar lo que se conoce como R^2 (coeficiente de determinación). En forma práctica, este valor representa en qué medida la variación de la variable dependiente es explicada por la variable independiente. Su valor también va de 0 a 1 y suele multiplicarse por 100 a fin de informar la proporción. Volvamos al ejemplo anterior: «el peso corporal se correlacionó en forma moderada con la edad ($r=0,79$, $p<0,03$)». En este caso, nuestra variable dependiente «peso corporal» muestra buena relación con la edad si calculamos R^2 de la correlación de Pearson es 0,62; al multiplicar este valor por 100 obtenemos 62 %, lo que refiere que un 62 % de la variación del peso corporal podría justificarse por el incremento de valor de edad.

Además de evaluar la relación entre variables cualitativas o cuantitativas, también se puede estudiar si existen diferencias de medias entre grupos. Este análisis se denomina t-Test para muestras independientes. Por ejemplo, si leemos: «los pacientes diabéticos presentaron mayor edad ($p=0,02$)», nos está diciendo que se comparó la edad (variable cuantitativa) entre personas con y sin diabetes (variable cualitativa dicotómica). En el caso de que la distribución de la variable cuantitativa no sea normal, se determina la diferencia de mediana con la prueba U de Mann-Whitney. Si la variable cualitativa presenta más de dos grupos, la diferencia de media se analiza mediante ANOVA de un factor (por su parte, la diferencia de mediana se realiza con el test Kruskal Wallis). Ambas pruebas suelen acompañarse de análisis *post-hoc* con el propósito de conocer no solo si hay diferencia entre grupos sino también entre qué grupos existe la diferencia.

Hasta este punto hemos descrito los análisis estadísticos comúnmente empleados para valorar asociación entre dos variables, sin embargo, también es factible determinar la influencia de una variable independiente (o más de una) respecto de un resultado particular (variable dependiente), esto es, se puede determinar el grado de predicción de variables independientes en cuanto a la variable dependiente.

Cuantificación de la asociación entre variables

Hemos mencionado que para analizar la relación entre dos variables cuantitativas se puede emplear el test de correlación de Pearson (o Spearman, en el caso de las que no se distribuyan con normalidad) y el coeficiente de determinación. Si, además de valorar el grado de asociación, nuestra intención es observar la influencia (grado de predicción) de la variable independiente sobre la variable dependiente (respuesta), la prueba estadística

que se emplea es la regresión lineal simple. Existe la posibilidad de que sea más de una la variable independiente que influye en el comportamiento de la variable dependiente y/o haya interacción entre las variables independientes. Esto se analiza mediante la regresión lineal múltiple. Es importante recordar que tanto las variables independientes como la dependiente deben ser cuantitativas.

Cuando la variable dependiente es cualitativa dicotómica (solo tiene dos categorías), la prueba estadística que se usa es la regresión logística; si tiene más de dos categorías, se emplea la regresión logística multinomial. En concreto, la regresión logística se realiza cuando se desea predecir la ausencia o presencia de una determinada condición; por ejemplo, conocer cuál es la chance (*odd ratio* —OR—) de sufrir un infarto agudo de miocardio de tener un índice de masa corporal categorizado como sobrepeso.

Las pruebas mencionadas, tanto lineal como logística, permiten confeccionar modelos de predicción que son usados habitualmente en la práctica clínica. Cada vez que determinamos *scores* para saber si nuestro paciente tiene riesgo de esto o aquello estamos haciendo uso de modelos predictivos.

Otras formas en las que habitualmente se detalla cuánto influye una o más variable en el *end-point* (resultado) que se investiga es el riesgo relativo o riesgo absoluto si el trabajo es prospectivo, mientras que en el caso de que el estudio sea retrospectivo se emplea *odd ratio*. El concepto de riesgo refiere a la presencia de un factor que incrementa la probabilidad de que se desarrolle un evento. Cuando hablamos de riesgo absoluto (RA) nos limitamos al riesgo que tiene determinada muestra (la expuesta o no expuesta a un factor determinado) de sufrir el evento en cierto tiempo. Por su parte, riesgo relativo (RR) mide la fuerza de asociación entre el factor y el evento y determina la probabilidad de que se desarrolle el evento comparando los grupos expuestos y no expuestos al factor. Observemos la siguiente tabla de contingencia.

TABLA 6. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DE RIESGO

	con ACV	sin ACV	total
Aspirina	20 (a)	980 (b)	1000
Sin aspirina	40 (c)	960 (d)	1000
Total	a+c	b+d	a+b+c+d

ACV: accidente cerebro-vascular.

Aquí, el factor es el tratamiento con aspirina y el evento es el desarrollo de accidente cerebrovascular. Mencionamos previamente que RA es cuando

valoramos el riesgo de los expuestos o no expuestos de desarrollar el evento, por lo tanto:

- RA de ACV en expuestos (medicados con aspirina) = $a/a+b$.
Es el cociente del número de personas que desarrollaron ACV estando medicadas con aspirina con relación al total de los expuestos (todos los medicados con aspirina), siendo $20/1000 = 0,02$.
- RA de ACV en NO expuestos (NO medicados con aspirina) = $c/c+d$.
Es el cociente del número de personas que desarrollaron ACV estando medicadas con aspirina con relación al total de los expuestos (todos los medicados con aspirina), siendo $40/1000 = 0,04$.

El RR es la fuerza de asociación entre el evento y la exposición al factor. Su cálculo se realiza comparando el RA de los expuestos respecto de los no expuestos.

- $RR = RAE/RANe$, es decir $(a/a+b) / (c/c+d) = 0,02/0,04$. El RR es 0,5
Para interpretar este valor del RR es importante saber que si:
 - $RR=1$, no hay asociación entre el factor y el desarrollo del evento.
 - $RR<1$, indica que la asociación es negativa, esto es, que el factor es «protector» en cuanto al desarrollo del evento.
 - $RR>1$, refiere que hay una asociación positiva entre el factor y el desarrollo del evento, por lo cual su presencia incrementa la frecuencia de que suceda el evento.

En nuestro caso, si el RR es de 0,5 (<1) estaría indicando que el tratamiento con aspirina es «protector» con relación al desarrollo del ACV. Al principio del capítulo definimos el término intervalo de confianza (IC) como el rango que contiene al valor observado, empleándose habitualmente el IC del 95 %. Cuando leemos los valores de «riesgo relativo» es muy importante observar el IC 95 %, porque si el rango incluye al valor 1, quiere decir que el RR puede ser tanto «protector» como «factor de riesgo». Volvamos al RR de nuestro ejemplo. El mismo es 0,5 y el IC 95 % es 0,2 – 0,8; como el rango del RR sigue siendo menor a 1, nuestro RR es «protector» respecto del desarrollo del evento. Ahora, si tenemos $RR=0,5$ IC 95 % 0,2 – 1,18, el margen del IC abarca al 1 e incluso lo supera, por lo que nuestro RR es a la misma vez «protector» y «factor de riesgo» para el mismo evento.

Otros parámetros que se pueden calcular sobre la base del riesgo relativo y absoluto son:

- Reducción de riesgo relativo: $(1-RR) \times 100$. Consiste en establecer en qué proporción se reduce el riesgo de sufrir el evento en presencia del factor con respecto al riesgo de la población no expuesta.
- Reducción de riesgo absoluto (RAR). Es la diferencia entre los grupos de RA no expuestos y RA expuestos ($RANe - RAE$). Mediante la misma se

pretende comparar generalmente tratamientos. En nuestro ejemplo, están el $RANe=0,04$ y $RAe=0,02$, por lo cual RAR es $0,02$. Esto significa que la administración de aspirina ha reducido el riesgo de ACV en un 2 %, es decir que la aspirina evita 2 ACV cada 100 pacientes tratados.

- Número Necesario a Tratar (NNT): $1/RAR$. Es un cálculo mediante el cual se determina cuántas personas requieren del tratamiento para evitar el desarrollo del evento. Por ejemplo, ante nuestra situación, RAR es de $0,02$; entonces el cálculo de NNT es $(1/0,02)$. El NNT es 50, esto es, se requiere tratar a 50 personas con aspirina para que 1 no desarrolle el evento.

Otra forma de cuantificar la asociación entre variables es mediante el OR, cuya traducción al castellano puede ser «razón de probabilidades» o «razón de momios». Se la considera otra forma de expresar el riesgo y representa el cociente entre *Odd* de expuestos y *Odd* de los no expuestos al factor. Es decir, el número de veces que se desarrolla el evento frente a cuántas veces no ocurre del grupo de los expuestos al factor respecto de la misma tasa pero de los no expuestos. Volvamos a nuestra tabla de contingencia de ejemplo.

TABLA 7. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DE OR

	con ACV	sin ACV	total
aspirina	20 (a)	980 (b)	1000
sin aspirina	40 (c)	960 (d)	1000
total	a+c	b+d	a+b+c+d

En este caso el *Odd* de los expuestos se determina a/b , mientras que el *Odd* de los no expuestos es c/d ; por lo cual OR es $(a/b)/(c/d)$. Acorde a los valores de la tabla, el OR es $0,5$. La interpretación del resultado de OR similar al de RR, puesto que, si $OR=1$, no hay asociación entre el factor y el evento; $OR<1$, la asociación es negativa, y $OR>1$ implica que la presencia del factor se asocia a mayor tasa de incidencia del evento.

Como podemos observar, tanto RR y OR nos dieron el mismo valor, sin embargo esto no siempre es así, solo sucede si la incidencia del evento de estudio es menor al 10 % o cuando la diferencia de la tasa de eventos entre grupos es muy pequeña.

VALIDACIÓN DE PRUEBAS DIAGNÓSTICAS

En este último apartado nos referiremos a las pruebas estadísticas que se emplean para determinar la eficacia de una prueba diagnóstica como

herramienta complementaria para el diagnóstico de una entidad nosológica. Entre las características principales de las pruebas diagnósticas se destacan:

- Exactitud: implica la correspondencia de la prueba con el valor verdadero.
- Precisión: reproductibilidad de la prueba cuando se repite en la misma muestra.
- Contraste con el estándar de oro: cuando una nueva prueba diagnóstica se va a emplear como estudio complementario para una determinada entidad nosológica, antes debe compararse con la prueba diagnóstica determinada como estándar de oro, esto es, el test diagnóstico de mayor eficacia para el diagnóstico de la patología en cuestión. Por ejemplo, considerando que la prueba estándar para realizar diagnóstico de fibrosis hepática es la biopsia, las pruebas diagnósticas que surjan con intención de ser empleadas para el diagnóstico de fibrosis hepática deben compararse con la biopsia.

En tanto, validar la prueba diagnóstica supone determinar algunos aspectos que mencionaremos a continuación, pero antes, a fin de ejemplificar, utilizaremos la siguiente tabla de contingencia:

TABLA 8. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DE VALIDACIÓN DE PRUEBA DIAGNÓSTICA

	Prueba estándar de oro positiva	Prueba estándar de oro negativa
Prueba En Estudio +	Vp	Fp
Prueba En Estudio -	Fn	Vn

VP: verdadero positivo, FP: falso positivo, VN: verdadero negativo, FN: falso negativo.

- Sensibilidad: efectividad de la prueba para identificar a las personas que tienen la condición en estudio $\rightarrow VP/VP+FN$.
- Especificidad: es la calidad de la prueba para determinar las personas verdaderamente sanas (sin la condición) $\rightarrow VN/FP+VN$.
- Valor predictivo positivo: nos dice qué probabilidad real tiene el sujeto de estar enfermo si la prueba de estudio le dio positiva $\rightarrow VP/VP+FP$.
- Valor predictivo negativo: en este caso, si el resultado de la prueba en estudio fue negativa, mediante la determinación del valor predictivo negativo, se calcula la probabilidad de que el sujeto realmente no esté enfermo $\rightarrow VN/FN+VN$.
- Likelihood ratio (LR): es el cociente de probabilidad de obtener determinado resultado mediante la prueba en estudio, es decir cuánto se modifica mi

pretest, pudiendo ser LR+ en caso de presentar un test positivo, o LR- si el resultado es negativo. Se calculan de la siguiente forma:

LR+: Sensibilidad/(1-Especificidad)

LR-: (1-Sensibilidad)/Especificidad

Pretest se puede definir como la probabilidad de que una persona presente una determinada condición dependiendo de la prevalencia de la misma en la población, las características propias de la persona (edad, género), manifestaciones clínicas, etc. El médico establece ante el motivo de consulta del paciente y el contexto del mismo presunciones diagnósticas en cuanto a la probabilidad de que la persona tenga o no las entidades nosológicas sospechadas por el profesional de la salud. Este pretest se modificará (aumentará o disminuirá) acorde al LR+/LR- del resultado del estudio complementario, determinándose la probabilidad posttest.

Existe una herramienta denominada Nomograma de Fagan que, conociendo el LR de la prueba complementaria y a partir del pretest de nuestra presunción diagnóstica, se puede calcular el posttest. Esto lo veremos más adelante en el capítulo de pruebas diagnósticas.

Hemos descrito en este capítulo los test estadísticos más empleados en el desarrollo de la MBE. En las próximas secciones iremos contextualizando cada uno de ellos en los diversos tipos de estudios científicos a fin de que pueda observar su aplicabilidad.

7 Reporte de caso y serie de casos

Caso clínico

Roberto, de 30 años, ingresa a internación con diagnóstico de neumonía por Influenza H1N1. Completado el tratamiento con Oseltamivir y por buena evolución recibe el alta hospitalaria y continúa con controles por consultorio externo. A la semana asiste a la consulta, refiere que le cuesta conciliar el sueño y que tiene alucinaciones auditivas. Ante esto, el profesional se pregunta si se ha descrito la presencia de manifestaciones psiquiátricas en personas que padecieron Gripe H1N1.

GENERALIDADES

Como se aprecia en el caso clínico planteado, la pregunta clínica se orienta a conocer si se encuentra descrito en la bibliografía científica situaciones similares a la planteada. Cabe remarcar que no se pregunta si existe asociación entre las manifestaciones explicitadas por el paciente y su antecedente de gripe A o si hay una relación causal entre el virus H1N1 y alteraciones psiconeurológicas.

En este marco, el tipo de trabajo que interesa para responder explícitamente la pregunta clínica es el reporte de caso o la serie de casos. El primero es la comunicación de una situación con una condición inusual en su forma de presentación, de la modalidad diagnóstica de la entidad nosológica o del tratamiento. Por su parte, el artículo científico denominado «serie de casos» hace referencia a un *paper* en el cual se incluye un grupo de pacientes que tienen como factor común la condición inusual. En ninguno de los dos tipos de trabajo se realiza contraste de hipótesis, por el contrario, son netamente descriptivos, detallándose en forma de historia la presentación clínica, la metodología diagnóstica, el tratamiento y la evolución. Consta, luego de presentarse el caso, de una breve discusión en la que se remarca el interés científico que implica lo «atípico» de la situación problema, generalmente mediante el desarrollo de una revisión sobre el tema.

Tanto el reporte de caso como la serie de casos comprenden la base de la pirámide de niveles de evidencia científica y es importante recordar que

no es correcto realizar inferencias, deducciones o confirmaciones mediante los mismos. Como son descripciones de situaciones particulares, no son generalizables y las «asociaciones teóricas» o la explicación de «causa» son planteos hipotéticos que deben ser contrastados adecuadamente mediante otros estudios. Sin embargo, ambos son útiles ya que nos permiten comunicar una «novedad científica» y promover el desarrollo de nuevas investigaciones, en particular cuando se relacionan con nuevas entidades nosológicas posibles, formas de presentación clínica, diagnósticas o de evolución, o efectos colaterales de fármacos y usos potenciales (beneficiosos o adversos) de los mismos.

Volvamos al caso clínico planteado al inicio de este capítulo. La pregunta clínica es saber si hay algún caso similar. Si hacemos la búsqueda en PubMed, empleamos como palabras clave «psychiatric symptoms», «H1N1» y «Oseltamivir», y seleccionamos la opción «case report» en tipo de artículo, podemos encontrar el artículo «Psychiatric Symptoms in a Patient with Influenza A (H1N1) Treated with Oseltamivir (Tamiflu): A Case Report», publicado en *Clinical psychopharmacology and neuroscience* en el año 2015 por Jeon Sang Won. En el mismo se describe la presencia de alucinaciones auditivas, sentimientos suicidas, deterioro cognitivo e insomnio en un hombre de 22 años luego de haber recibido tratamiento con Oseltamivir. El artículo no señala una asociación entre el tratamiento y las manifestaciones, postula la posibilidad de que exista una relación entre el fármaco y el desarrollo de manifestaciones psiquiátricas, por lo cual sería conveniente explorar si, por algún mecanismo particular, Oseltamivir puede llegar a inducir síntomas psiquiátricos.

En resumen, el reporte de caso y la serie de casos son trabajos descriptivos útiles para comunicar hechos no conocidos y promover el desarrollo de investigaciones científicas.

8 Estudios transversales o transaccionales

Caso clínico

Jorge tiene 47 años y antecedente de hipertensión arterial, medicado con Enalapril 10 mg/día. Asiste a la consulta con los resultados de laboratorio solicitados para su control periódico y en el mismo se informa ácido úrico e índice aterogénico elevados.

GENERALIDADES

La pregunta que se puede formalizar, en el contexto de la situación planteada, es: ¿existe relación entre los niveles de ácido úrico y el índice aterogénico plasmático? La misma ya nos establece que nuestro objetivo de búsqueda es encontrar estudios que valoren asociación entre variables, por lo cual queda descartado el primer nivel de evidencia, reporte de caso o la serie de casos. En esta oportunidad, cuando nos limitamos a valorar simplemente correlación o asociación el tipo de estudio a buscar es el transversal (también denominado transaccional).

Los trabajos transversales son aquellos que no presentan: a) secuencia temporal, b) cronología de los hechos respecto del inicio del estudio, y c) intervención por parte del investigador. Es decir que son trabajos observacionales sin continuidad en el eje tiempo, lo que imposibilita conocer causalidad. Las ventajas de este tipo de estudio comprenden su sencillez de realización, su bajo costo y la fácil obtención de resultados (asociaciones/correlaciones).

Pueden plantearse desde un enfoque descriptivo o analítico. En el primer caso, su meta es brindar detalles sobre situaciones epidemiológicas y estimar magnitud y distribución de un evento. Por ejemplo, si quisiéramos valorar el número de personas con antecedente de enolismo atendido en el servicio de clínica médica de un hospital durante el año 2014, estaremos pautando un objetivo descriptivo en un margen de tiempo establecido y cuya metodología sería determinar la frecuencia (magnitud) de personas asistidas con el antecedente de enolismo (evento de interés).

Si el diseño del trabajo es analítico, su fin es analizar en forma simultánea la exposición al factor (variable independiente) y el evento de interés (variable

dependiente) en una población (muestra) en un momento determinado. Los análisis estadísticos que se emplean son de asociación o de correlación, como por ejemplo χ^2 o r de Pearson, respectivamente. Recordemos que al no ser una serie temporal y al no permitir establecer la cronología de los hechos (exposición–evento) con relación al inicio del estudio los resultados no permiten identificar relaciones causales.

ANÁLISIS E INTERPRETACIÓN DE LA ASOCIACIÓN

Retomemos del Capítulo 5, las tablas en las cuales se ejemplifican las pruebas estadísticas en cuanto a las variables de estudio.

TABLA 9. ANÁLISIS ESTADÍSTICOS EMPLEADOS PARA CONTRASTAR LA HIPÓTESIS EN MUESTRAS NO RELACIONADAS

	Cuantitativas	Cualitativas
Cuantitativas	Pearson (r) – R^2 Rho de Spearman	2 Grupos: T-Test U M. Whitney (NP)
Cualitativas	3 o más grupos: ANOVA Kruskal Wallis (NP)	χ^2 (Chi-cuadrado) Exacta de Fisher (NP)

NP: no paramétricas.

TABLA 10. ANÁLISIS ESTADÍSTICOS EMPLEADOS PARA CONTRASTAR LA HIPÓTESIS EN MUESTRAS RELACIONADAS

	2 mediciones	3 o más mediciones
Cuantitativas	T-Test datos apareados Prueba de Wilcoxon (NP)	Análisis de la Varianza Análisis de la Varianza de Friedman (NP)
Cualitativas	Prueba de McNemar	Q de Cochran

NP: no paramétricas.

Según cuál fuera el objetivo del trabajo de investigación, son varias las pruebas estadísticas que se emplean. Por ejemplo, podemos leer en un texto científico:

- «Se conformaron acorde al grosor del tejido adiposo epicárdico dos grupos. Aquellos con mayor grosor presentaron menor tasa de filtrado glomerular ($p=0,02$)». En este caso se ha comparado el valor de la tasa de filtrado glomerular entre dos grupos (comparación de variable cuantitativa con respecto a una variable cualitativa).

TABLA 11. RELACIÓN DEL TEJIDO ADIPOSO EPICÁRDICO CON EL FILTRADO GLOMERULAR

	Con grosor de tejido adiposo epicárdico normal	Con grosor de tejido adiposo epicárdico aumentado	p
Tasa de filtrado glomerular (media±DS)	91±13 mL/min/ 1,73 m ²	72±7 mL/min/ 1,73 m ²	0,02

- Si ambas variables continuas, grosor adiposo de tejido epicárdico y tasa de filtrado glomerular, se categorizaran en variables dicotómicas (dos grupos), se podría terminar asociación mediante X^2 (Chi-cuadrado) o Exacto de Fisher. Supongamos que el estudio incluyó 100 personas.

TABLA 12. ASOCIACIÓN DEL NÚMERO DE PERSONAS CON FILTRADO GLOMERULAR DISMINUIDO Y CON TEJIDO EPICÁRDICO AUMENTADO

	Con grosor de tejido adiposo epicárdico normal	Con grosor de tejido adiposo epicárdico aumentado
Tasa de filtrado glomerular normal	10	8
Tasa de filtrado glomerular disminuida	9	73

- Como se puede apreciar en la tabla de contingencia, queda claro que la distribución por filtrado glomerular es diferente entre grupos. En este caso, el resultado es $X^2=16,2$; $p=0,0001$. Refleja la asociación entre las variables tasa de filtrado glomerular disminuida con grosor de tejido adiposo epicárdico aumentado.
- Si lo que se observa es relación entre las variables cuantitativas, se puede describir correlación de Pearson r (o su equivalente no paramétrico Spearman), «el incremento del grosor del tejido adiposo epicárdico se correlacionó significativamente con disminución del filtrado glomerular ($r= - 0,21$, $p=0,002$)».

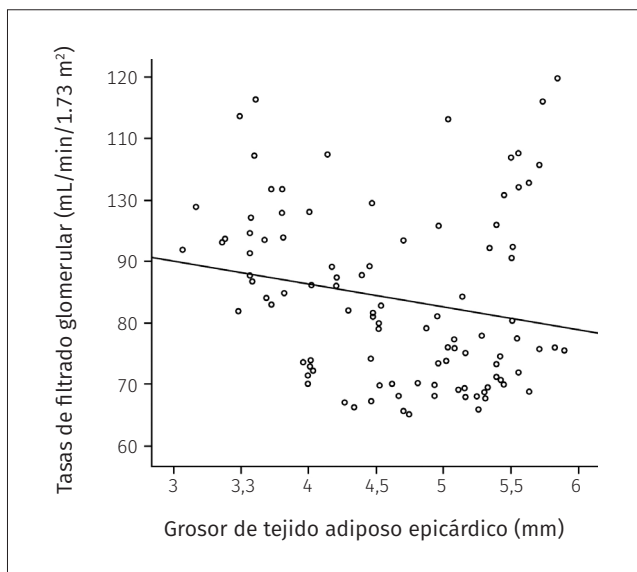


FIGURA 7. RELACIÓN ENTRE EL INCREMENTO DEL TEJIDO ADIPOSO EPICÁRDICO Y LA DISMINUCIÓN DEL FILTRADO GLOMERULAR

Con independencia de cuál fuera la prueba estadística empleada, lo único que nos refieren los investigadores es que observaron que las personas con mayor medida del grosor del tejido adiposo epicárdico presentaron menor filtrado glomerular; esto no implica que el incremento del grosor del tejido adiposo epicárdico sea causa o factor de riesgo de disminución del filtrado glomerular. Sin embargo, al presentarse dicha correlación se genera un nuevo planteo hipotético: ¿el grosor del tejido adiposo epicárdico influye en la tasa de filtrado glomerular? Por lo tanto, las conclusiones observacionales analíticas pueden ser disparadores de nuevos trabajos científicos.

LIMITANTES DE LOS TRABAJOS TRANSVERSALES

Dentro de las limitantes de los estudios transversales se destacan principalmente dos. Por un lado, no permiten definir la asociación encontrada al ámbito individual. Esto obedece al hecho de que el estudio se desarrolla con información resumida, no individual. Volviendo al ejemplo anterior de la correlación: «el incremento del grosor del tejido adiposo epicárdico se correlacionó significativamente con disminución del filtrado glomerular ($r=0,63$, $p=0,002$)», estamos analizando factor y evento de interés en un grupo sin considerar la variabilidad individual.

La otra limitante es lo que se denomina «falacia ecológica» y se debe a la imposibilidad por parte del investigador de controlar eventuales variables confundidoras en el análisis. Esto puede suceder porque las pruebas estadísticas de asociación o correlación no contemplan la medición de covariables (variables independientes que pueden modificar la variable independiente de interés o la variable dependiente en estudio), o este tipo de información puede no haber sido incluida al pautarse el protocolo de trabajo (sesgo de inclusión de datos). Mucho palabrerío que puede sonar imposible de entender. Vamos a tratar de esquematizarlo en el ejemplo mencionado de la relación entre el grosor del tejido adiposo epicárdico y el filtrado glomerular: «el incremento del grosor del tejido adiposo epicárdico se correlacionó significativamente con disminución del filtrado glomerular ($r=0,63$, $p=0,002$)». Primero identifiquemos cada parte del ejemplo nuevamente:

- Nuestro evento de interés es la disminución del filtrado glomerular (variable dependiente).
- Nuestro factor de interés es el grosor del tejido adiposo epicárdico (variable independiente).
- Covariables: serían todos los factores que pueden modificar tanto el filtrado glomerular o el grosor del tejido epicárdico.

Como se puede observar en el ejemplo, no se mencionan covariables, solo la relación entre las variables independiente y dependiente. Pero si se supiera que el antecedente de hipertensión arterial de larga evolución promueve el incremento del grosor del tejido epicárdico, la variable «hipertensión arterial» podría jugar un rol confundidor en la correlación observada.

Resumidas las características del estudio transversal, procedamos a analizar la situación problema con la cual abrimos el actual capítulo. Primero establezcamos la pregunta clínica, que podría ser: ¿existe relación entre el nivel sérico de ácido úrico y el del índice aterogénico? Con las palabras clave «uric acid», «atherogenic index of plasma», con el filtro de búsqueda: en título y abstract, se puede encontrar el artículo «Association of uric acid, atherogenic index of plasma and albuminuria in diabetes mellitus», publicado en International Journal of Clinical and Experimental Medicine en el año 2014. En este artículo se observa correlación positiva entre ácido úrico y el índice aterogénico plasmático ($r=0,154$, $p<0,001$).

9 Estudios de casos y controles

Caso clínico:

Alberto, de 53 años, con antecedentes de hipertensión arterial en tratamiento con Enalapril e Hidroclorotiazida 10/12,5 mg por día, consulta por monoartitis. Tras el examen físico y estudios complementarios pertinentes se le realiza el diagnóstico de gota. Ante esta situación, Alberto pregunta si puede ser que las medicaciones que está tomando para el control de la presión influyeran en el desarrollo de su actual patología.

GENERALIDADES

En este caso, la pregunta guía que puede plantearse es si los fármacos empleados para el tratamiento de hipertensión arterial favorecen el desarrollo de gota: «¿será que el evento que tiene el paciente se debe a que presenta determinado antecedente?». Al pautarse la cronología de los eventos, existen dos posibles tipos de estudio que nos pueden dar respuestas, los trabajos cohortes o el de casos y controles. Sin embargo, en esta oportunidad se enfoca al planteo desde la presencia del evento, lo cual implica que hay que mirar desde el mismo hacia atrás a fin de analizar si los antecedentes del paciente se relacionaron con el desarrollo del evento en cuestión. El tipo de estudio de interés, entonces, es el caso y control, el cual pertenece a los estudios epidemiológicos analíticos observacionales retrospectivos.

Como su nombre lo indica, en su diseño se contemplan dos grupos, uno denominado «casos», constituido por los sujetos que presentan el evento de interés (variable dependiente que se quiera estudiar), y el otro llamado «control», que refiere a las personas sanas. Una vez conformados los grupos, se explora y compara en forma retrospectiva la exposición relativa a diferentes factores (variables independientes) que podrían tener un rol predisponente para el desarrollo de la condición de interés.

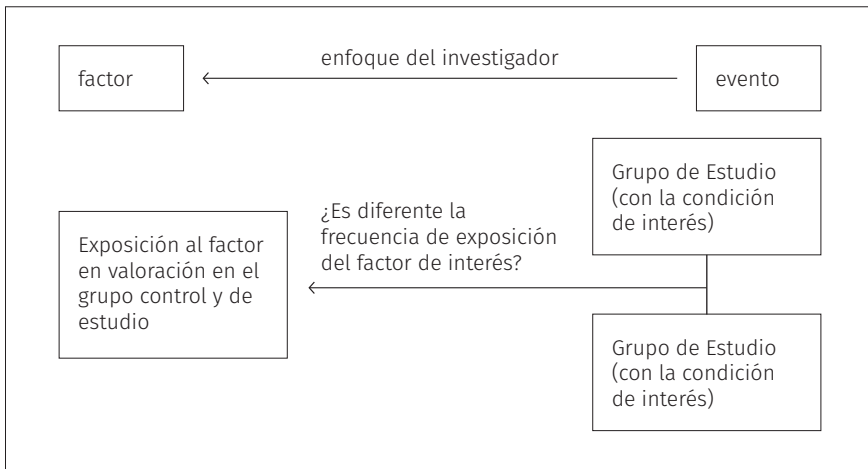


FIGURA 8. ESQUEMA DEL DESARROLLO DE LOS TRABAJOS CASOS Y CONTROLES

TABLA 13. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DEL OR

	Con la condición de interés	Sin la condición de interés	Total
Expuestos al factor	A	B	A+B
No expuestos al factor	C	D	C+D
Total	A+C	B+D	A+B+C+D

El objetivo del estudio es analizar la probabilidad de padecer el evento respecto de la exposición al factor, es decir el *odd* de expuestos, de no expuestos y la razón de probabilidades entre grupos (OR). De esta forma, se calcula si la proporción de los casos expuestos es mayor que la de los controles. Recordemos que en los estudios de caso y control no se determina riesgo relativo sino que, mediante OR, se analiza el grado de asociación entre el o los factores (variables independientes) y la presencia de la condición (variable dependiente). En circunstancias específicas, el valor de OR puede ser similar al riesgo relativo, como estimador de riesgo (incidencia del evento de estudio es menor al 10 %, o cuando la diferencia de la tasa de eventos entre grupos es muy pequeña).

Los estudios casos y controles son de mucha utilidad para analizar problemas de baja frecuencia o cuando se requiere un abordaje de una situación en particular en forma relativamente rápida. Otras ventajas que presentan es que requieren de un tamaño muestral menor que el de los estudios cohorte y evitan los inconvenientes propios de estos. Por ejemplo, cuando el desarrollo de las enfermedades es muy prolongado, la estrategia de cohorte implicaría

mucho más dinero y tiempo. En cuanto a las desventajas, considerando que el diseño es retrospectivo, no se puede estimar la incidencia de la condición o el tiempo entre la exposición al factor y la aparición del efecto. Es importante remarcar que, si bien el enfoque de los trabajos casos y controles es retrospectivo, puede suceder que sujetos desarrollen el evento con posterioridad al inicio del estudio, por lo que se incluirían en el grupo de casos en forma prospectiva hasta cierto tiempo pautado con anterioridad.

CONFORMACIÓN DE LOS GRUPOS

Los criterios de selección que se empleen siempre van a afectar los resultados y, por lo tanto, la conclusión del estudio. Es por ello que los mismos deben estar debidamente especificados en el texto científico. Definir la situación de caso establece los parámetros que delimitan las características del grupo control.

Por ser de diseño retrospectivo, la obtención de los casos surge generalmente de las historias clínicas o registros médicos. Más allá de la fuente de los datos, el grupo de estudio debe de representar la historia natural de la condición de interés y todos los portadores de esta deben de tener igual probabilidad de ser seleccionados para participar del estudio.

Por su parte, el grupo control o de referencia se define por las mismas condiciones que los casos (con la diferencia de que estos últimos presentan la condición de interés) y deben pertenecer a la misma población de la cual provienen los casos. Otro hecho importante para la selección de controles es que tienen que contar con la misma probabilidad de exposición al factor que los sujetos del grupo de casos, y que sea similar al grupo de casos en lo que respecta a otras variables que pudieran ser confundidoras en el desarrollo de la condición de interés. Una estrategia empleada para cumplir con este último requisito es el apareamiento de las características de los grupos (las proporciones de características de los grupos son similares) o del individuo (para cada caso se selecciona un control similar). Respecto del número de personas a incluir como control, por lo general se realiza en una proporción 1:1, 1 control por cada caso. Sin embargo, incrementar el tamaño muestral del grupo control aumenta el poder de los resultados, en especial en situaciones en las cuales el número de casos es bajo.

Existen tres condiciones a considerar al realizar la definición de la muestra a incluir en los grupos: el grado de severidad de la condición de interés, el nivel de exposición al factor en estudio y el lugar de selección. La primera sucede cuando la enfermedad o evento que se va a analizar en su evolución natural presenta niveles de gravedad; si se incluyera un solo grado, el grupo no será representativo de todas las personas con la condición. En la segunda,

puede ocurrir que la exposición al factor no sea homogénea tanto en el grupo de casos como en el de control, incluso las etapas de la evolución natural de la condición pueden estar relacionadas con el grado de exposición al factor en estudio. Si hay diferentes niveles de exposición al factor se debe analizar la relación entre las variables independientes y la dependiente en forma estratificada por exposición al factor. Es decir, la tabla de contingencia habitual quedaría conformada de la siguiente forma:

TABLA 14. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DEL *ODD RATIO* ANTE INTENSIDAD VARIABLE DE EXPOSICIÓN AL FACTOR

	Con la condición de interés	Sin la condición de interés	Total
Fuertemente expuestos al factor	A	B	A+B
Levemente expuestos al factor	C	D	C+D
No expuestos al factor	E	F	E+F
Total	A+C+E	B+D+F	A+B+C+D+E+F

La tercera condición que se debe tener en cuenta para la definición de las muestras es el lugar de selección. Cuando los casos pertenecen a un solo lugar de selección, por ejemplo, un hospital, los sujetos incluidos en el estudio pueden tener características únicas que no tienen otros individuos con la misma condición. Esto limita luego la extrapolación de resultados.

Por último, cuando se lee un estudio de casos y controles debemos identificar los sesgos que puede presentar su diseño; recordemos que sesgo se denomina al error no aleatorio. Dentro de este tipo de trabajo se encuentran los sesgos de selección de sujetos y obtención de la información. Una forma de evitar ambos es mediante el diseño de caso y control anidado en una cohorte, que permite que el investigador conozca en detalle las características de los sujetos en cuanto a exposición (grado, tiempo), variables confundidoras y desarrollo del evento.

ANÁLISIS E INTERPRETACIÓN DE ASOCIACIÓN

Las pruebas estadísticas de asociación bivariada o correlación, así como diferencias de medias entre dos o más grupos, son test que pueden emplearse en los diseños de casos y controles, pero el objetivo central de este tipo de estudio es comparar las proporciones del evento de interés respecto de la exposición o no a un factor. Esto se realiza mediante la determinación del OR.

Los estudios de casos y controles no permiten establecer la razón de riesgos o la razón de tasas de incidencias. Estos son pruebas estadísticas propias de los trabajos cohortes en los cuales la muestra al inicio del estudio no presenta el evento, por lo que admite estimar el número de sujetos que desarrolla la condición de interés. Sin embargo, sí se puede calcular la razón de momios, y es un buen estimador de la asociación entre la exposición al factor en estudio y la condición de interés (ver Capítulo 6).

La razón de momios u OR se estima mediante el cociente entre el *odd* de los expuestos (a/b) y el *odd* de los no expuestos (c/d). Si la razón es igual a 1, entonces no existe asociación entre la/s variable/s independiente/s y la condición de interés; si es <1, implica que la exposición al factor en estudio se asocia de manera inversa con el evento o, dicho de otro modo, disminuye la posibilidad de desarrollar la condición. Finalmente, si es >1, la asociación es positiva, por lo que el factor en estudio favorece el desarrollo del evento.

TABLA 15. TABLA DE CONTINGENCIA ENTRE EXPOSICIÓN A FACTOR Y CONDICIÓN DE INTERÉS

	Con la condición de interés	Sin la condición de interés	Total
Expuestos al factor	A	B	A+B
No expuestos al factor	C	D	C+D
Total	A+C	B+D	A+B+C+D

- *Odd* expuestos = a/b
- *Odd* no expuestos = c/d
- OR = (a/b) / (c/d)

Recordemos que, a la par del resultado del cociente, debe estar expresado el intervalo de confianza, que es el rango que contiene al valor observado, ya que, si bien el OR puede ser mayor a 1 el IC 95 % (intervalo de confianza del 95 %), puede incluirlo y entonces carece de valor real el OR. Por ejemplo, si leemos en un trabajo científico: «los IECAS en personas con hipertensión arterial podrían relacionarse con el desarrollo de gota (OR 2,1, IC 95 % 0,8-3,4)», esta afirmación es errónea debido a que el intervalo de confianza no solo abarca al 1 (si OR=1 no existe asociación) sino que también es <1 (si OR<1 la exposición al factor disminuye la posibilidad de desarrollar el evento), y el mismo factor no puede ser protector, no estar relacionado y ser perjudicial a la misma vez.

Cuando el valor de OR es mayor a 1 (y el intervalo de confianza no lo incluye) se puede determinar lo que se denomina el riesgo atribuible poblacional y de los expuestos (fracción etiológica), que es el impacto de la exposición al

factor con respecto al desarrollo del evento, o la proporción de la condición que se evitaría si los sujetos no estuvieran expuestos al factor en estudio. Su valor se calcula de la siguiente forma:

- Proporción de riesgo atribuible poblacional $\rightarrow [pe (OR-1)] / [pe (OR-1)+1] \times 100$
- pe = proporción de expuestos, OR = *odd ratio*.
- Fracción etiológica en los expuestos $\rightarrow (OR - 1)/OR$

En los artículos científicos es común el uso del análisis estadístico regresión logística para determinar el OR . Este test permite incluir más de una variable al análisis. Si observamos la tabla de contingencia, la misma es de doble entrada, expuesto o no expuesto a un factor y presencia o ausencia del evento; solo admite analizar la exposición al factor que el investigador determine. Pero puede suceder que exista más de una variable independiente (factor de exposición en estudio) asociada al desarrollo de la condición o coexistan variables confusoras, es decir, variables que pueden influir en el resultado de la asociación entre el factor y el evento pero que no estén directamente relacionadas con el desarrollo de la condición. En concreto, el test regresión logística posibilita disminuir el sesgo de confusión y analizar múltiples factores de exposición.

PAUTAS PARA LEER UN TRABAJO CASOS Y CONTROLES

- ¿Cómo fue la inclusión de los sujetos de los grupos de casos y controles?
- ¿Es homogéneo el grado de exposición al factor entre e intergrupos?
- ¿Los sujetos del grupo de casos presentan la misma etapa de la evolución natural de la condición de interés?
- ¿El grupo control fue conformado de manera apareada con el grupo de casos? De ser así, ¿el apareo fue individual o por características del grupo?
- ¿El factor de interés existió siempre con antelación al evento?
- ¿Hay variables confundidoras? Si es así, ¿se incluyeron en el análisis estadístico?
- ¿Las medidas de asociación están descritas en forma adecuada?
- ¿La conclusión del trabajo es coherente con los resultados de los análisis estadísticos?

Resumidas las características del estudio Casos y Controles, analicemos la situación problema con la cual abrimos el actual capítulo. Mencionamos que la pregunta clínica podría ser: ¿los fármacos empleados para el tratamiento de hipertensión arterial favorecen el desarrollo de gota? Considerando que en el caso ya se explicitan dos fármacos, IECAS y Tiazidas, podríamos buscar

en forma independiente cada una de ellas o realizar una primera búsqueda en forma generalizada.

Usando las palabras clave «Antihypertensive drugs», «gout», «hipertensión», con el filtro de búsqueda en cuanto al tipo de estudio: *case and control*, se puede encontrar el artículo «Antihypertensive drugs and risk of incident gout among patients with hypertension: population based case-control study», publicado en *British Medical Journal* en el año 2012. Dentro de las conclusiones del artículo se menciona que el uso de IECA se asoció a la presencia de gota y la probabilidad de riesgo es de 3,17, con un intervalo de confianza del 95 % de 2,90 a 3,47. Como se aprecia, el riesgo es mayor a 1 y el IC no incluye este valor, por lo cual el tratamiento de IECA en personas hipertensas es un factor vinculado al desarrollo de gota.

El texto citado menciona otros fármacos empleados para el tratamiento de hipertensión arterial que también se encuentran asociados al desarrollo de gota. Lo invitamos a buscar el artículo, de acceso gratuito, para poner en práctica lo aprendido en este capítulo.

10 Estudios de cohorte

Caso clínico

Alberto, con antecedente de hipertensión arterial y diagnóstico reciente de gota, asiste a la consulta con los resultados de control de los niveles de ácido úrico, los cuales persisten elevados. Preocupado por su situación, Alberto buscó en Internet que el ácido úrico, al encontrarse incrementado, podría causar infartos. Ante la incertidumbre de que esto fuera verdad, le consulta a su médico si es así.

GENERALIDADES

Hemos mencionado en el capítulo anterior dentro de las desventajas de los estudios, casos y controles que los mismos no permiten determinar tasa de incidencia de eventos y causalidad. Por el contrario, el diseño cohorte admite calcular ambos como consecuencia de las características de temporalidad y seguimiento de los sujetos. En los trabajos cohortes, el investigador se posiciona desde la exposición al factor en estudio y valora su asociación con el desarrollo del evento (riesgo relativo, riesgo absoluto). En concreto, los estudios de cohorte son observacionales para valorar la relación entre factores de riesgo potenciales y el desarrollo de una condición en una población en un tiempo determinado.

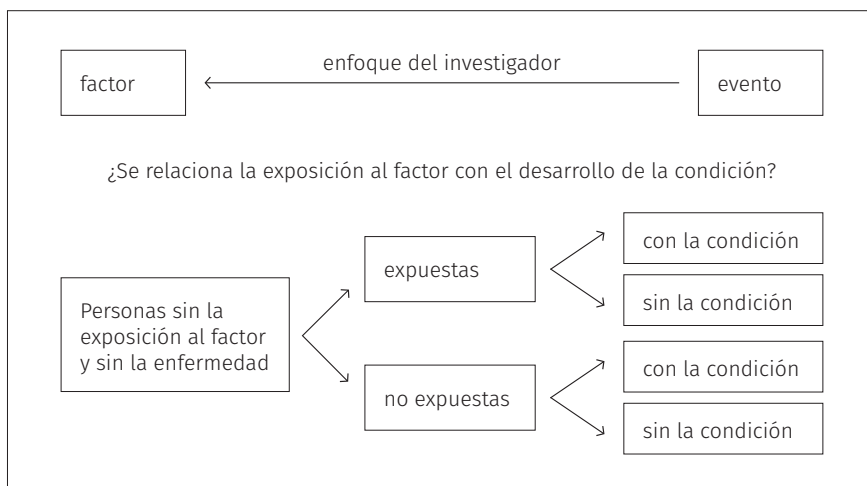


FIGURA 9. ESQUEMA DEL DESARROLLO DE LOS TRABAJOS COHORTE

Sobre la base del caso clínico planteado, en el cual Alberto desea conocer si el nivel de ácido úrico en sangre es un factor de riesgo cardiovascular, la pregunta clínica estaría orientada a saber si la exposición a un factor (incremento del ácido úrico) se relaciona con el desarrollo de una condición (evento cardiovascular). El enfoque puede ubicarse en el diseño de cohorte. Eso garantiza que la exposición precede a la ocurrencia del evento, e incluso permite analizar la asociación entre múltiples exposiciones y efectos (cuando la muestra se conforma con anterioridad a la exposición).

Existen dos grandes grupos de estudios cohorte:

- **Prospectivos:** cuando las personas incluidas en el trabajo no presentan el evento de interés (por ejemplo, enfermedad cardiovascular) y se analiza la exposición a factores de riesgo que podrían influir en el desarrollo del mismo en un tiempo determinado. Para esto se identifica a los sujetos expuestos o no al factor en el presente y, a partir de ese punto de referencia en el tiempo, se los sigue por un período especificado.
- **Retrospectivo:** si bien la metodología es similar, el estudio se concreta en forma *post-hoc*, es decir que se identifica a los individuos que estuvieron expuestos y no expuestos al factor en algún momento en el pasado y se los sigue hasta observar la ocurrencia del evento en la actualidad (o en el pasado reciente).

Llevar a cabo estudios de cohorte implica sortear diversas limitaciones, como el costo económico, el tiempo de seguimiento, la selección adecuada de los grupos de expuestos (diversidad en los grados de exposición), el hecho de mantener el tamaño muestral durante el seguimiento y los cambios en los criterios y métodos diagnósticos que pueden sucederse durante el proceso.

No obstante, sus ventajas son mayores y hacen que estos estudios tengan un valor importante dentro de la construcción del saber en las ciencias médicas en la comprobación de asociaciones causales e identificación de riesgo.

CONFORMACIÓN DE LOS GRUPOS

Al momento de la inclusión de los sujetos en el estudio deben tener posibilidades de desarrollar el evento y el tamaño muestral tiene que ser representativo de la población. Los criterios de selección, vinculados a los objetivos del estudio, deben ser los mismos tanto para el grupo de expuestos como para el de comparación. Es preferible que exista apareamiento de los individuos o de los grupos a fin de disminuir las chances de que variables confundidoras intervengan en la asociación entre la exposición y el desarrollo del evento.

El grupo de control puede ser interno o externo. En el primer caso es cuando se selecciona una muestra de una población y luego se clasifican los grupos en cuanto a la presencia o ausencia de la exposición al factor. Por su parte, la comparación externa se da al seleccionar los individuos expuestos de la población y, con base en las características del grupo, escoger los sujetos no expuestos al factor de la misma población.

Por último, mencionamos antes que uno de los problemas a evitar cuando se desarrolla un estudio de cohorte es sostener el número de personas incluidas en el trabajo a lo largo del seguimiento. Es esperable que un porcentaje se pierda, y esto tiene que ser contemplado a la hora de pautar el tamaño muestral del estudio. Al menos un 95 % del total debe continuar en el seguimiento o, en caso contrario, los resultados que se obtengan perderán validez.

EVALUACIÓN E INTERPRETACIÓN DE LA ASOCIACIÓN

De igual forma que los estudios transversales y casos y controles, la presentación de los resultados de los trabajos de cohorte se va a realizar mediante estadística descriptiva y analítica. Incluso, tal como los diseños de casos y controles, los estudios de cohorte permiten determinar el OR (ver Capítulos 5 y 8). Pero la principal relevancia estadística que presenta este diseño es que admite calcular (además de incidencia) riesgo.

Retomemos los conceptos de riesgo que fueron explicados en el Capítulo 5:

- Riesgo Absoluto (RA): riesgo que tiene determinada muestra (la expuesta o no expuesta a un factor) de sufrir el evento en cierto tiempo.
RA de presentar el efecto entre los expuestos (RAE) $\rightarrow a/(a+b)$.
RA de presentar el efecto entre los no expuestos (RANE) $\rightarrow c/(c+d)$.

- Riesgo Relativo (RR): mide la fuerza de asociación entre el factor y el evento y establece la probabilidad de que se desarrolle el evento comparando los grupos expuestos y no expuestos al factor $\rightarrow RAE/RANE$.
Si el resultado del cociente es igual a 1, entonces no existe asociación entre la/s variable/s independiente/s y la condición de interés; si es <1 implica que la exposición al factor en estudio se asocia de manera inversa con el evento o, dicho de otro modo, disminuye la posibilidad de desarrollar la condición. Finalmente, si es >1 , la asociación es positiva, por lo que el factor en estudio favorece el desarrollo del evento. Recordemos la importancia del IC para la interpretación de este valor.
- Riesgo atribuible: es el impacto de la exposición al factor con relación al desarrollo del evento, la proporción de incidencia del evento en los sujetos expuestos como consecuencia del factor en estudio $\rightarrow (a/a+b) - (c/c+d)$. La determinación del riesgo atribuible porcentual o fracción etiológica se hace de la siguiente forma: $[(a/a+b) - (c/c+d) \times 100] / (a/a+b)$.
- Riesgo atribuible poblacional: es la diferencia de tasas de incidencia del evento en la población general y la población no expuesta al factor de riesgo y expresa cuánto de la incidencia total de la condición es debido a la exposición del factor $\rightarrow (a+c)/(a+b+c+d) - (c/c+d)$.
El riesgo atribuible poblacional porcentual se calcula mediante la fórmula: $[(a+c)/(a+b+c+d) - (c/c+d) \times 100] / [(a+c)/(a+b+c+d)]$.

PAUTAS PARA LEER UN ESTUDIO DE COHORTES

- ¿Cuál es el objetivo del estudio de cohortes?
- ¿Qué tipo de estudio de cohorte es?
- ¿Cómo fue la selección de los grupos de estudio y control?
- ¿Existió apareamiento de grupo o individuo?
- Respecto del seguimiento de los pacientes, ¿fue adecuado para la comprobación de la relación causa-efecto?
- ¿Hay variables confundidoras? Si es así, ¿se incluyeron en el análisis estadístico?
- ¿Las medidas de asociación están descritas en forma adecuada?
- ¿La conclusión del trabajo es coherente con los resultados de los análisis estadísticos?

Resumidas las características del estudio de cohortes, procedamos a realizar el análisis de la situación problema. En la misma se planteaba el caso de Alberto, quien deseaba saber si el nivel incrementado de ácido úrico sérico es un factor de riesgo cardiovascular. Esquematicemos entonces la pregunta

clínica mediante la metodología PICO pero tengamos en cuenta que no comparamos métodos, por lo cual «c» (comparación) puede omitirse.

- Paciente: hombres de mediana edad (*middle-aged men*).
- Intervención: determinación del nivel de ácido úrico (*uric acid level*).
- *Outcome* (resultado): factor de riesgo cardiovascular (*risk factor for cardiovascular disease*).

Si empleamos la estrategia de búsqueda con las palabras en inglés y con el filtro de trabajos de cohortes, dentro de los resultados podemos rescatar el artículo «Uric Acid Level as a Risk Factor for Cardiovascular and All-Cause Mortality in Middle-aged Men», publicado en *Archives of Internal Medicine* en el año 2004. Este artículo científico describe la asociación entre el factor (incremento del ácido úrico sérico) y el desarrollo de muerte cardiovascular con una media de seguimiento de 12 años en hombres adultos, siendo el riesgo relativo de 3,73 (ic 95 % 1,42-9,83). Es decir que el incremento de ácido úrico aumenta la probabilidad, en hombres, de padecer un evento cardiovascular como un infarto agudo de miocardio.

11 Análisis de sobrevida

Caso clínico

Pese a las medidas preventivas y de control de los factores de riesgo cardiovascular, Alberto, quien padece de hipertensión arterial y gota, sufre un infarto agudo de miocardio, motivo por el cual ingresa a internación. Su médico, ante esta situación clínica, se pregunta si el nivel de ácido úrico se relaciona con infartos letales o no.

GENERALIDADES

En el marco de la situación clínica planteada, se establece que el uso de uno u otro fármaco puede influir en la sobrevida del paciente. Uno de los primordiales objetivos de los estudios, en especial de los ensayos clínicos, es valorar la efectividad de un tratamiento para reducir la mortalidad mediante el análisis de sobrevida. Referirnos al análisis de supervivencia es hacer mención a la expresión cuantitativa del pronóstico de una entidad nosológica o problema de salud, es decir, de su evolución natural. Es importante remarcar que en muchas ocasiones las pruebas estadísticas empleadas para este fin son utilizadas para determinar frecuencia de eventos ligados a la historia natural del problema salud en cuestión.

Cuando un trabajo científico contempla llevar a cabo el análisis de supervivencia se debe tener en cuenta que el factor de exposición es el diagnóstico de la entidad nosológica o intervención que recibiera el sujeto como tratamiento farmacológico. En tanto, el evento de interés es el resultado de la exposición, como mortalidad-sobrevida u otro evento vinculado al factor. Tomemos como ejemplo el artículo científico citado en el capítulo anterior y que es de utilidad para responder la pregunta que se plantea el médico en el caso clínico disparador: «Uric Acid Level as a Risk Factor for Cardiovascular and All-Cause Mortality in Middle-aged Men». En ese escrito se hace el seguimiento de hombres durante un promedio de 12 años aproximadamente y se valoran los eventos que padecieron los sujetos y su relación con el nivel de ácido úrico sérico elevado, entre ellos, muerte por causas cardíacas o *stroke*.

Las características más importantes a tener en cuenta al interpretar el análisis de supervivencia son el tiempo de seguimiento, la modalidad de identificación de la condición (enfermedad o problema de salud) y la determinación del resultado que se espera medir.

TIEMPO DE SEGUIMIENTO

Como se puede apreciar, este tipo de estudio se estructura en el diseño de cohorte (sea retrospectivo o prospectivo), mediante el cual se delimita un período de observación. La escala de tiempo debe estar definida de manera apropiada, al igual que el evento de interés. Puede suceder que la duración del seguimiento sea distinta para los sujetos incluidos en el estudio, esto se debe a que, si el mismo es abierto, las personas pueden incorporarse en momentos temporales distintos.

Uno de los inconvenientes que podemos encontrar al leer un artículo en el cual se valore sobrevida es la censura de datos. Esto sucede cuando el seguimiento termina antes de que se cumpla el evento en estudio, y existen tres motivos que pueden generar la censura:

- a) El sujeto se pierde durante el seguimiento y no hay información sobre el mismo.
- b) El paciente abandona el estudio.
- c) Se suspende el estudio antes de que aparezcan los eventos.

Otro acontecimiento que limita la interpretación adecuada de la información es el truncamiento de los datos. Esto ocurre de dos modos: uno cuando se encuentra la inclusión tardía del sujeto al estudio y el comienzo con posterioridad a la exposición al factor el tiempo de observación. Y el otro modo es cuando solo se incluyen los individuos que presentan el evento.

A fin de poder obtener un buen análisis de sobrevida se debe esperar que los datos no censurados o truncados representen adecuadamente a aquellos que por alguna de las tres razones mencionadas fueron censurados o truncados.

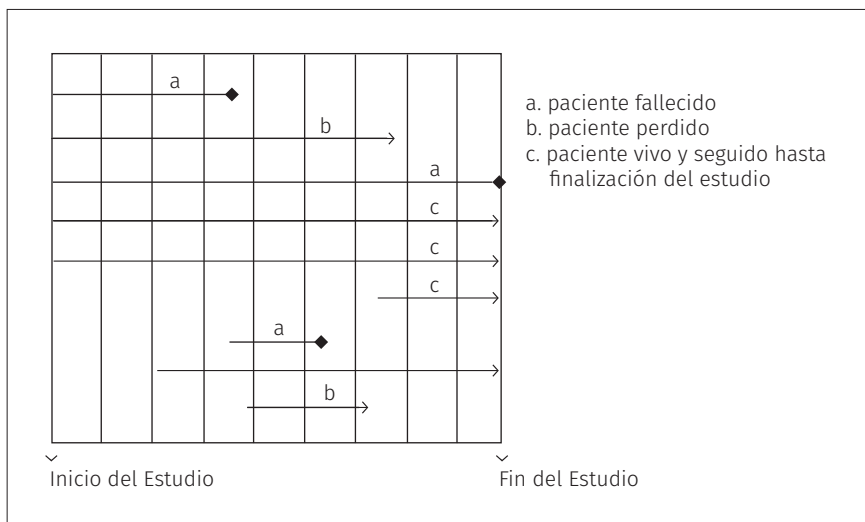


FIGURA 10. ESQUEMA DE PLANILLA DE SEGUIMIENTO PARA DETERMINAR SOBREVIVENCIA

MÉTODOS BÁSICOS DE ANÁLISIS DE SOBREVIVENCIA

Hay varias pruebas estadísticas para analizar la supervivencia o el evento de interés luego de la exposición al factor en cuestión. Una forma sencilla es la determinación de la tasa de letalidad.

$$\text{Letalidad (\%)} = \frac{\text{Nº de muertes por una enfermedad en un periodo de tiempo} \times 100}{\text{Nº de casos diagnosticados de la misma enfermedad en el mismo periodo}}$$

$$\text{Mortalidad (\%)} = \frac{\text{Nº de muertes por una enfermedad} \times 1000}{\text{Población expuesta al riesgo}}$$

FIGURA 11. FÓRMULAS PARA EL CÁLCULO DE LA TASA DE LETALIDAD Y DE MORTALIDAD

Por lo general se la emplea ante enfermedades agudas y no para aquellas de larga duración y determina la probabilidad de morir para las personas que tienen esas enfermedades. La tasa de letalidad no es lo mismo que la tasa de mortalidad; esta última es la proporción de gente que fallece respecto del total de la población.

Por su parte, la supervivencia a 5 años (tiempo establecido en forma arbitraria), de uso muy frecuente en la evaluación del abordaje diagnóstico y terapéutico en enfermedades neoplásicas, es la proporción de individuos que luego de 5 años de seguimiento se encuentra viva con relación al total de la muestra a la que se le realizó el diagnóstico y/o inició determinada intervención.

$$\text{Sobrevivida a 5 años (\%)} = \frac{\text{Nº de sujetos vivos a los 5 años desde el diagnóstico o inicio del tratamiento de la enfermedad}}{\text{Nº de personas diagnosticadas o que iniciaron el tratamiento de la enfermedad}} \times 100$$

FIGURA 12. FÓRMULA PARA EL CÁLCULO DE SOBREVIVIDA A 5 AÑOS

TABLA DE VIDA

La tabla de vida (o mortalidad) es una presentación tabular del progreso del seguimiento de una cohorte o de un subgrupo de la misma. En esencia, calcula la probabilidad de sobrevivir en segmentos de tiempo y asume que no hay cambios en la intervención durante el tiempo de observación. A su vez, cada caso contribuye al análisis sin considerarse el segmento de tiempo en el cual fue incluido. Existen dos formas de tabla de vida: por un lado, la tabla actuarial, que aporta un enfoque transversal del evento mortalidad o supervivida durante un período corto en general no mayor al año. Por otro lado, la tabla de cohorte, que consiste en el seguimiento desde el inicio del evento de análisis hasta que concluya el período de observación o se produzca el fallecimiento del sujeto.

Vamos a analizar una tabla de supervivencia de enfoque cohorte a fin de apreciar cómo se utiliza por medio de un estudio hipotético de pacientes tratados entre 2010 y 2014 y seguidos hasta 2015 (sin pérdidas).

TABLA 16. TABLA DE SUPERVIVENCIA, PRIMERA ETAPA

Año	Tratados	Número de sobrevivientes al año de la detección				
		2011	2012	2013	2014	2015
2010	213	34	29	21	17	11
2011	87		15	13	31	15
2012	73			16	29	21
2013	45				14	6
2014	78					59

Ahora reordenemos la tabla con referencia al tiempo transcurrido desde el año de inicio y el tiempo de sobrevida con el propósito de determinar la probabilidad de sobrevivir por año.

TABLA 17. TABLA DE SUPERVIVENCIA, SEGUNDA ETAPA

Año	Tratados	Número de sobrevivientes al año de la detección				
		1er año	2do año	3er año	4to año	5to año
2010	213	34	29	21	17	11
2011	87	15	13	31	15	
2012	73	16	29	21		
2013	45	14	6			
2014	78	59				
Totales	496	138	76	73	32	11

Para calcular la probabilidad de sobrevida al 1º año, hacemos $138/495=0,28$. Ya para el 2º año debemos sustraer al total del primer año aquellos sujetos que pertenecen al último período de análisis (2014 a 2015) y proceder al cálculo de la siguiente forma $76/(138-59) \rightarrow 76/79 \rightarrow 0,96$. Lo invitamos a continuar calculando las probabilidades de supervivencia por año; debe quedar de la siguiente forma:

TABLA 18. TABLA DE SUPERVIVENCIA, TERCERA ETAPA

Probabilidad de sobrevida				
1er año	2do año	3er año	4to año	5to año
0,28	0,96	1,04	0,61	0,64

Determinada la supervivencia por año, podemos obtener la probabilidad de sobrevida acumulada mediante la multiplicación de la probabilidad de sobrevida por año. En nuestro ejemplo, el resultado es 0,1. Si queremos amplificar el valor para una mejor lectura, se multiplica el mismo por 100. De esta forma, la probabilidad acumulada de sobrevivir postratamiento sería del 10 % en 5 años.

CURVA DE SOBREVIDA

Otra modalidad para analizar supervivencia es mediante curvas de supervivencia utilizando la curva de Kaplan Meier. Es importante notar que se trata de un método no paramétrico, por lo cual no asume función de probabilidad. Si bien en el capítulo sobre bioestadística mencionábamos la preferencia por el uso de pruebas estadísticas paramétricas, en el caso de análisis de supervivencia no es así. Existen test paramétricos para valorar supervivencia que demandan ajustar los datos al modelo de análisis ya que la presencia del evento no tiene una distribución normal. Por eso, en la práctica, es preferible el uso de estadísticos no paramétricos para estudiar supervivencia.

La curva de Kaplan Meier es usada para determinar la proporción de supervivencia sin fijar segmentos de tiempo como en las tablas de vida; por el contrario, es el evento lo que define el período de seguimiento del sujeto. Esto es, el paciente es seguido desde su inclusión hasta que sufre el evento de interés, por ejemplo, su fallecimiento. Por su parte, los individuos «perdidos» son removidos del análisis. La curva se grafica al unir el final de período de cada sujeto que ha padecido el evento excluyendo los perdidos, como se puede ver en la siguiente Figura (líneas punteadas).

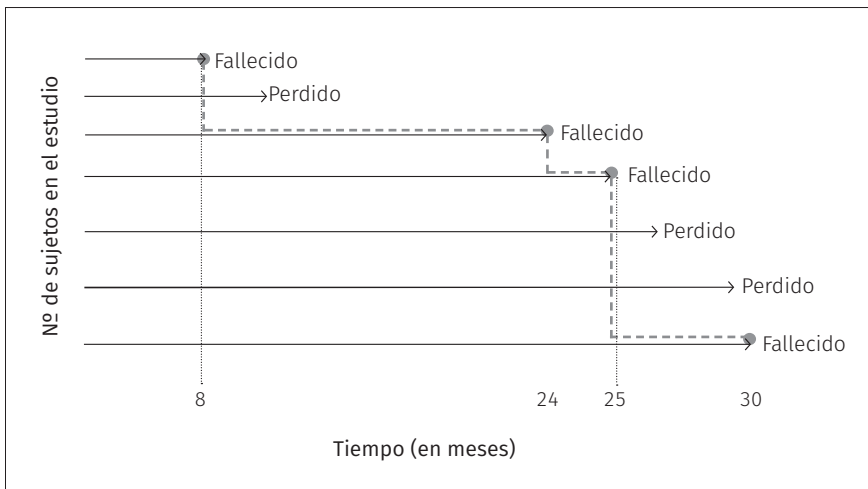


FIGURA 13. CURVA DE SOBREVIDA

Lo representado allí se extrapola a una tabla mediante la cual podemos determinar la proporción acumulada que sobrevivió al final del seguimiento.

En los ensayos clínicos suele realizarse la comparación de curvas de Kaplan Meier entre las distintas ramas de estudio, por ejemplo, entre el grupo placebo y el de tratamiento en estudio. Es una forma de comprobar la eficacia de la nueva intervención en comparación con el tratamiento estandarizado para una entidad nosológica.

Y para concluir este capítulo, mencionaremos la regresión de Cox. Esta prueba estadística ha adquirido relevancia en los últimos años como instrumento de análisis de sobrevida. Al igual que la regresión logística, que hemos comentado con anterioridad, la regresión de Cox busca determinar las variables independientes que se relacionan e influyen en la variación de riesgo y, en definitiva, también sobre la función de sobrevida de los individuos del estudio respecto del evento de interés. En concreto, es una prueba estadística de regresión donde se analizan factores que modifican el riesgo de padecer un daño. El evento de interés no necesariamente debe de ser mortalidad, puede ser, por ejemplo, infarto agudo de miocardio o cualquier complicación propia de la enfermedad de base y/o su abordaje.

La regresión de Cox debe entenderse como la regresión logística, y como tal también su resultado es una tasa denominada *Hazard Ratio* y su interpretación es similar al OR o RR.

Si recurrimos nuevamente a la lectura del artículo «Uric Acid Level as a Risk Factor for Cardiovascular and All-Cause Mortality in Middle-aged Men» veremos en él curvas de Kaplan Meier en las cuales se compara, con relación al nivel de ácido úrico sérico, la sobrevida acumulada. Acompañando estas gráficas se encuentra el análisis estadístico correspondiente de regresión de Cox cuyo resultado arroja que valores entre 5,05 y 5,88 mg/dL de ácido úrico presentan un *hazard ratio* de mortalidad por causas cardiovasculares de 2,7 (IC 95 % 1,20 a 6,04). Esto quiere decir que los niveles del ácido úrico entre 5,05 y 5,88 incrementan el riesgo de muerte por causas cardiovasculares.

Hasta aquí comentamos en forma resumida la mayoría de las modalidades de evaluación de sobrevida y en el próximo capítulo, que trata sobre ensayo clínico, veremos estudios en los cuales es común que se comuniquen análisis de supervivencia, entre otras cosas.

TABLA 19. TABLA DE PROPORCIÓN ACUMULADA DE SOBREVIVENCIA

Tiempo hasta la muerte desde el inicio del tratamiento	Nº vivo en cada momento	Nº fallecido en cada momento	Proporción que falleció en ese momento	Proporción que sobrevivió en ese momento	Proporción acumulada que sobrevivió en ese momento (supervivencia acumulada)
8	7	1	0,14	0,86	0,86
24	5	1	0,2	0,8	0,68
25	4	1	0,25	0,75	0,51
30	1	1	1	0	0

12 Ensayos clínicos

As to different methods of treatment, it is possible for us to assure ourselves of the superiority of one or other (...) by enquiring if the greater number of individuals have been cured by one means than another.
(Pierre Louis)

Caso clínico

Luego de padecer el infarto, Alberto presenta disminución de la fracción de eyección del ventrículo izquierdo, siendo menor al 35 %. Ante esto, su médico sabe que la espirolactona está indicada, pero desconoce si la misma disminuye la mortalidad en personas con insuficiencia cardíaca severa.

GENERALIDADES

En el siglo XIX el médico francés Pierre Charles Alexandre Louis interesado en comprobar la real eficacia de la práctica médica, introdujo el método numérico mediante el cual promovía cuantificar y comparar las respuestas de los pacientes a los diversos tratamientos que recibían para una misma entidad nosológica. De esta forma, cuestionó la indicación de la sangría como medida terapéutica en la neumonía, demostrando su ineficacia. Este fue uno de los primeros ensayos clínicos realizados empleando la estadística como herramienta para contrastar una hipótesis, con los déficits metodológicos propios de la época.

Como su nombre lo indica, los estudios de ensayos clínicos son trabajos científicos de enfoque prospectivo longitudinal, en los cuales el investigador realiza la experimentación controlada en voluntarios humanos de tratamientos o intervenciones contra problemas de salud de cualquier tipo a fin de valorar fundamentalmente su eficacia y seguridad. El objetivo de mayor relevancia que persiguen los estudios clínicos es el de demostrar que la nueva intervención reduce la mortalidad de las personas incluidas en el

ensayo. Sin embargo el mejorar la calidad de vida de los sujetos también es un propósito que como profesionales de la salud no podemos dejar de lado.

Las principales diferencias con los estudios observacionales son que en primer lugar no cuentan con los criterios de selección de «presencia de evento» (caso y control) o «exposición» independiente del investigador (cohorte), y por otro lado, el investigador participa activamente asignando a un grupo la exposición al factor de interés (exposición dependiente del investigador).

Otro rasgo de los ensayos clínicos, es que toda condición del estudio es controlado con el propósito de que cada etapa del mismo y accionar de investigadores sea uniforme para evitar posibles errores que generen cuestionamientos a las conclusiones obtenidas.

Existen diversos diseños de ensayos clínicos, pero en concreto y en base a los objetivos del estudio, la muestra poblacional representativa de la población blanco que participará del mismo se ha de conformar por individuos que compartan características similares sociodemográficas y de la condición de la enfermedad. Posteriormente, en forma aleatoria sin intervención de ningún miembro que participa del estudio, los sujetos se asignarán a un grupo que va a estar sujeto a la intervención en estudio, y a un grupo de control. Por un tiempo determinado se realizará el seguimiento correspondiente obteniendo los datos de interés y realizando tanto el análisis general y estratificado en subgrupos por variables como edad, sexo, gravedad de la enfermedad, entre otros.

FASES DE UN ENSAYO CLÍNICO

Antes de comenzar con los diseños y análisis de los ensayos clínicos, vamos a realizar un breve comentario sobre el desarrollo de nuevas intervenciones. El foco del tratamiento no debe de centrarse solamente en la eficacia, sino que su seguridad ha de ser lo principal a valorarse.

La primera etapa es la investigación preclínica que se realiza en modelos experimentales (animales) y se evalúan las características farmacodinámicas, farmacocinéticas y seguridad de la nueva droga. Posteriormente se da inicio a la fase I del ensayo terapéutico en la cual se estudian los efectos valorados con anterioridad en un grupo reducido de personas (farmacología humana). Durante la fase 2 se evalúa el efecto terapéutico en un número bajo de individuos que presentan la entidad nosológica de interés, determinándose la dosis óptima, intervalo de administración y tasa de eventos relacionadas a su uso. En base a los resultados obtenidos se da inicio a la 3ª fase en la cual la investigación se orienta a comprobar el efecto terapéutico (y en algunas ocasiones comparar su eficacia en relación con otro fármaco). En este periodo también se reestablecen los efectos adversos.

Los ensayos clínicos que habitualmente lee el profesional de salud para saber si hace uso o no de una nueva intervención corresponden a la comunicación de los resultados de la 3ª fase. Sin embargo no es el último nivel del desarrollo farmacológico. Los protocolos de investigación se diseñan para realizar el seguimiento de los sujetos en forma estricta por un tiempo determinado, una vez concluido el mismo finaliza la 3ª fase. Una vez que comienza a comercializarse el fármaco, inicia la fase 4 que corresponde a un trabajo observacional en el cual se analiza la presencia de efectos adversos raros no apreciados durante las etapas anteriores, y a la relación costo-beneficio de su empleo.

En forma resumida estas son las etapas previstas para el desarrollo de una nueva intervención. En el presente capítulo abordaremos el diseño y análisis estadísticos usados habitualmente en las comunicaciones científicas de ensayos clínicos que corresponden a la tercera etapa. Por último, existe un registro de ensayos clínicos que permite conocer quienes lo realizan, la etapa en la cual se encuentra, y otros datos que hacen a la metodología del mismo. Dicha base es de acceso gratuito en línea mediante el sitio de internet www.clinicaltrials.gov.

DISEÑOS DE ENSAYOS CLÍNICOS EN RELACIÓN CON LA ASIGNACIÓN DE LOS GRUPOS

El investigador puede llevar a cabo el ensayo clínico no sólo para tratar de demostrar superioridad de una intervención sobre otra, sino también para comprobar equivalencia o no inferioridad entre ellas.

- El ensayo de superioridad tiene como fin detectar que una intervención es mejor que otra (generalmente un placebo). Es decir que el contraste de hipótesis es en base a la hipótesis nula de que no hay diferencias, por lo tanto al comparar las observaciones mientras menor sea el valor de p mayor será la probabilidad de que los resultados de la intervención en estudio se deban al tratamiento y no al azar.
- El diseño de los ensayos de equivalencia son para confirmar que no existe disparidad entre las intervenciones. Para analizar este objetivo se determina el margen de diferencia clínica que define el rango aceptable para realizar la comparación. Obtenidos los resultados se valora el intervalo de confianza de las diferencias del efecto entre las intervenciones. Si el valor del efecto con el respectivo intervalo de confianza para cada intervención cae dentro del rango establecido de equivalencia, se considerará que existe equivalencia entre los tratamientos.
- El estudio de no inferioridad se lleva a cabo para demostrar que un nuevo tratamiento no es menos efectivo que el que se emplea en la actualidad,

lo cual no implica que es mejor. A diferencia del diseño de equivalencia, en este caso sólo nos interesa que el valor mínimo del intervalo de confianza del efecto observado por la nueva intervención no sea menor al margen inferior establecido. Esto se llama intervalo de confianza de una sola cola.

Conocer estas formas de enfocar el contraste de las hipótesis es interesante para el profesional de la salud particularmente cuando los estudios son de equivalencia o de no inferioridad. Supongamos que el tratamiento actual para una determinada patología de mal pronóstico es económico y con escasos efectos adversos que de suceder no ponen en riesgo la vida; nuestro juicio de la relación costo beneficio nos dice que sería conveniente indicar el tratamiento. Ahora si en un ensayo clínico de equivalencia o de no inferioridad se demostró que una nueva intervención, para la misma enfermedad en cuestión, es igual o no inferior al tratamiento previo, sin embargo el costo y/o la tasa de efectos adversos es mayor, *¿cuál sería la relación costo-beneficio? ¿valdría la pena usar el nuevo fármaco?* Salvo que la intervención novel modifique claramente la relación costo-beneficio a favor de los beneficios en la persona con el problema de salud, no hay necesidad de reemplazar el fármaco.

Habiendo hecho mención de los enfoques de contraste de hipótesis que puede asumir el investigador, mencionaremos ahora las tres principales características metodológicas de los ensayos clínicos. En primer lugar, nos encontramos con el proceso de aleatorización de la asignación de los sujetos cuyo propósito es disminuir las diferencias entre los grupos, promoviendo de esta forma que la inferencia causal observada al concluir el estudio no sea consecuencia de otros factores. Es decir, siendo homogénea entre los grupos la distribución de variables que pudieran influir en el efecto de la intervención, de observarse diferencias deben de considerarse relacionadas al tratamiento en estudio. Existen diversas técnicas de aleatorización para la asignación las intervenciones entre las que se encuentran la simple, estratificada, en conglomerado o adaptativas a los tratamientos. La aplicación de una u otra depende del tipo de ensayo clínico que se realiza, pero independientemente cual fuera, el objetivo perseguido por la técnica de aleatorización continua siendo el mismo, disminuir la probabilidad de que el resultado se deba al azar.

Ahora, *¿qué haría Ud. si fuera un paciente que participa del estudio y sabe que recibe el tratamiento que se está probando?*, y en el caso de que fuera el médico tratante *¿cuál es su seguridad al indicar un medicamento que no sabe si realmente funciona?* Diversas cuestiones desde el aspecto personal del paciente, del médico tratante y/o del investigador pueden influir en el desarrollo del ensayo clínico si conocen que intervención están recibiendo o indicando. Por esta razón una de las condiciones ideales al llevarse a cabo el estudio es el cegamiento de sus participantes. Existen básicamente tres niveles:

- Simple ciego: uno de los tres miembros, habitualmente los sujetos incluidos en el estudio, desconoce la intervención asignada.
- Doble ciego: en estos casos ningún participante conoce los tratamientos que se aplican en las diversas ramas.
- Triple ciego: es el ensayo clínico en el cual los datos se encuentran blindados de análisis por un determinado periodo.

Por último, el otro punto a considerar en el diseño de los ensayos clínicos es el uso de placebo, agente farmacológico inactivo, como rama control de la nueva intervención. El mismo debe de contar con las mismas características y forma de administración al tratamiento en estudio. Si bien la finalidad del placebo es no inducir respuesta biológica alguna, se ha descrito ya hace más de 50 años que podrían promover mediante mecanismos inespecíficos efectos terapéuticos modificando el curso de la evolución natural del problema de salud. Hoy en día no es frecuente el uso de placebo en ensayos clínicos ya que de existir un tratamiento efectivo para la enfermedad sobre la cual se va investigar la nueva intervención, debe de aplicarse el mismo.

En conocimiento de las formas de contraste de hipótesis y características metodológicas principales de los ensayos clínicos, concluiremos este apartado con la descripción de los tipos de diseños de estudios clínicos en relación con las ramas de asignación de los grupos que habitualmente se emplean.

1. De diseño paralelo

Es el más simple de los diseños, en el cual los individuos son asignados en forma aleatoria a los respectivos grupos y permanecen en ellos hasta la finalización del estudio.

2. Diseño de tratamiento cruzado

Conformados los dos grupos, a un determinado tiempo se realiza el cruce entre ellos. De esta forma, aquellos que recibieron la intervención A pasan al grupo de intervención B y viceversa, siendo cada grupo entonces de estudio y control a la vez. Si bien la ventaja de este tipo de diseño es importante, existen varias limitantes propias de la condición del problema de salud, los individuos y la intervención, que imposibilitan llevarlo a cabo. Por ejemplo, si el tratamiento A modifica el curso de la enfermedad favorablemente a largo plazo sin volver al estado basal, no podría dilucidarse adecuadamente si la buena evolución bajo la intervención B se debe a ésta o al efecto aún del tratamiento A.

Un ejemplo de este tipo de diseño es el estudio denominado «*Morphine interaction with prasugrel: a double-blind, cross-over trial in healthy volunteers*» publicado en la revista *Clinical Research Cardiology* en 2015. Dicha investigación tenía como objetivo analizar si la morfina intervenía en la farmacocinética del prasugrel, considerando que disminuye la biodisponibilidad del clopidogrel. Se asignó entonces en forma randomizada 6 individuos sanos al grupo de intervención A (prasugrel 60 mg vía oral asociado a morfina 5 mg endovenoso) y 6 sujetos al grupo de intervención B (prasugrel 60 mg vía oral asociado a placebo —solución de cloruro de sodio 0.9 % endovenoso—). Luego de la administración de los fármacos, se esperó un periodo de 14 días considerado necesario para que las personas del estudio retornen al estado basal eliminando la posibilidad de efectos residuales («wash out»). Al día 15 los sujetos que fueron incluidos en el grupo A recibieron la intervención correspondiente al grupo B (prasugrel 60 mg vía oral asociado a placebo —solución de cloruro de sodio 0.9 % endovenoso—) y los del grupo B, la intervención del grupo A (prasugrel 60 mg vía oral asociado a morfina 5 mg endovenoso).

Como se puede apreciar existe el cruce de grupos en relación con la intervención en estudio. En el caso ejemplo, a través de la metodología empleada, se pudo observar que el uso de morfina en forma concomitante con el prasugrel disminuye la concentración máxima del metabolito activo de este último, sin embargo no afecta su efecto antiagregante en forma clínicamente relevante.

3. Diseño de tratamiento de reemplazo

En ocasiones nos encontramos ante problemas de salud que nos demandan modificar el esquema terapéutico sustituyendo un fármaco por otro. Los ensayos clínicos de diseño de tratamiento de reemplazo son empleados para estudiar los efectos que se producen al realizar este cambio de drogas; en una primera instancia los grupos son sometidos a la misma intervención (A) y luego de un determinado tiempo un grupo recibe el tratamiento B y el otro el C, comparándose luego los resultados observados.

4. Diseño factorial

Hemos mencionado anteriormente que el tamaño muestral del estudio debe de ser representativa de la población blanco, por lo tanto valorar dos o más intervenciones y sus combinaciones en un ensayo clínico de diseño paralelo implicaría incrementar la población a incluir en la investigación lo cual limitaría poder llevar a cabo este tipo de análisis. El diseño factorial surge como alternativa, ya que permite asignar en forma aleatoria las múltiples

intervenciones en forma independiente, y de esta forma evaluar el efecto de los diversos tratamientos y combinaciones.

El estudio clínico factorial 2 por 2 es el más simple y empleado; en el mismo los sujetos son asignados a una de las cuatro ramas posibles. Por ejemplo si tengo dos tratamientos, A y B, los grupos serían 1. intervención A, 2. intervención B, 3. intervención A y B, y 4. sin intervención (control).

PAUTAS PARA LEER UN ENSAYO CLÍNICO

Los ítems a considerar para una adecuada lectura de un ensayo clínico pueden resumirse en las siguientes preguntas que debe realizarse el lector para interpretar el estudio:

- ¿Cuál es el objetivo del ensayo?
- ¿El tamaño de la muestra es adecuado?
- ¿Las personas incluidas en el estudio se asignaron aleatoriamente?
- La realización del ensayo se hizo ¿simple, doble o triple ciego?
- Los grupos ¿eran similares al inicio del ensayo?
- Independientemente de la intervención ¿los grupos se trataron de igual forma?
- En el análisis de resultados ¿se incluyen a las personas que abandonaron el ensayo?
- ¿Durante qué tiempo se logró el efecto?
- ¿Está bien medido el efecto?
- ¿Es clínicamente significativo el efecto?

Al principio del capítulo mencionábamos que los ensayos clínicos principalmente buscaban demostrar la disminución de mortalidad o presencia de evento adverso de una enfermedad ante la aplicación de una nueva intervención. A su vez, como se está valorando un nuevo tratamiento también es importante que se describan los efectos adversos que puede inducir el mismo. Esto pauta el objetivo del trabajo de investigación y es muy importante porque el tamaño muestral se determina en relación con éste.

Consideremos el caso clínico disparador en el cual Luz de 59 presenta insuficiencia cardíaca severa y el profesional de salud sabe que estaría indicado el uso de espironolactona, pero no recuerda el motivo. Si la reducción de mortalidad es habitualmente el objetivo de los ensayos clínicos la pregunta clínica podría ser ¿la espironolactona reduce la mortalidad en pacientes con insuficiencia cardíaca severa? Si establecemos el método PICO tendríamos:

- Paciente: individuos con insuficiencia cardíaca severa (*patients with severe heart failure*).
- Intervención: espironolactona (*spironolactone*).

- Comparación: placebo (*placebo*).
- *Outcome* (resultado): reducción de la mortalidad (*reduced mortality*).

A través de esta metodología de búsqueda científica encontraremos el artículo denominado «*The effect of spironolactone on morbidity and mortality in patients with severe heart failure*» publicado en la revista *The New England Journal of Medicine* en el año 1999. El propósito del ensayo era demostrar que el uso de 25 mg de espironolactona en forma diaria disminuía la mortalidad en personas con insuficiencia cardíaca severa. Fue un ensayo doble ciego de diseño paralelo de asignación aleatoria entre los grupos de intervención A (placebo) y B (espironolactona). Lo invitamos a buscar y leer el artículo a medida que vamos avanzando en este apartado.

El objetivo del estudio fue analizar la reducción de la mortalidad ante el tratamiento con espironolactona, por lo cual el número necesario de personas a incluir se calculó considerando que la mortalidad en el grupo de estudio sería 17 % menor a la de control (errores $\alpha=0.05$ y $\beta=90$ %). Determinado el tamaño muestral, las intervenciones fueron asignadas aleatoriamente. La razón de la randomización es que las variables que pudieran interferir en el efecto biológico del tratamiento en estudio se distribuyan en forma homogénea entre los grupos. Esto se traduce generalmente en una tabla de comparación de las características de los grupos al inicio del ensayo en la cual se refiere si hubo diferencia de alguna de las variables. En el ensayo clínico que estamos usando de guía se puede apreciar que los grupos no presentaban diferencias entre ellos.

Un factor importante que influye en la interpretación de los resultados, es la forma en la cual los datos observados fueron analizados. Recordemos que el ensayo clínico es un trabajo prospectivo de seguimiento, por lo cual puede suceder que por varias razones los sujetos asignados a una rama del estudio no continúen en ella pasando a ser datos perdidos. Existen básicamente dos modalidades estadísticas, el análisis por protocolo y el método de análisis intención a tratar. En el primer caso se incluye en la evaluación estadística a aquellos sujetos que cumplieron con el protocolo de estudio, son los «casos válidos». Esta forma nos permite conocer el real efecto de la intervención en las personas que lo recibieron, sin embargo al excluirse a aquellas personas que no completaron adecuadamente el estudio, puede disminuirse la potencia del estudio, es decir, la capacidad de detectar la diferencia entre los tratamientos cuando esta diferencia es verdadera. Otra desventaja de este tipo de análisis es la pérdida de la distribución homogénea entre los grupos, promoviéndose que terceras variables puedan influir en el resultado final.

Por su parte, en el método de intención a tratar, todas las personas incluidas en el estudio son analizadas acorde a la asignación de la intervención manteniendo la comparabilidad entre los grupos. En otras palabras, este enfoque no tiene en cuenta las variaciones de protocolo, que los sujetos se retiren del

estudio, o cualquier otra cosa suceda luego de la primera asignación aleatoria. Ahora, ¿por qué se usa si no considera las modificaciones del estudio? Pues bien, este escenario es el más parecido a la práctica clínica real en el cual no todos los pacientes cumplen eficazmente con el tratamiento; y por otro lado, evita estimaciones demasiado optimistas sobre la eficacia de una intervención que puede suceder de no analizar los datos de las personas que se retiraron del estudio. Sin embargo, a su vez, en los ensayos de equivalencia o de no inferioridad el análisis de intención a tratar puede favorecer que no se observe diferencias entre las intervenciones.

En el ensayo clínico de guía para nuestro caso el método empleado fue el de intención a tratar. A su vez, al momento de calcular el tamaño muestral se incluyó un porcentaje extra de sujetos al grupo de tratamiento con espirolactona debido a que se tuvo en cuenta que era factible que un número de personas no completara adecuadamente el protocolo. Esto es algo habitual al momento del diseño de ensayos clínicos.

Finalmente, el último punto a considerar es en relación con el efecto estudiado, la forma en que fue evaluado, el tiempo en la que se logra el efecto y si clínicamente es significativo. En el ensayo sobre el beneficio de la espirolactona en reducir la mortalidad en personas con insuficiencia cardíaca severa, la investigación se suspendió antes de tiempo al objetivarse una reducción importante de la tasa de defunción en aquellas personas bajo tratamiento con el fármaco. Éste era el objetivo principal y para lo cual fue diseñado el estudio. Es importante siempre cotejar si los resultados que comunica el investigador están relacionados al objetivo primario del ensayo clínico, ya que el tamaño muestral y diseño del mismo se enmarcan centrados en el mismo. Si los investigadores, independientemente del resultado de contraste de hipótesis, comunican observaciones secundarias pero que no comprenden al objetivo principal, o mejor dicho, para lo cual no estaba diseñado el estudio, se debe tener en cuenta que la validez estadística no es la misma.

Por su parte la significancia clínica, hace referencia a la relevancia que tiene la observación en la práctica profesional. Si en un ensayo clínico se concluye que una droga es superior a otra en forma significativa estadísticamente, no implica que por sí mismo la diferencia sea importante desde el aspecto clínico. Un ejemplo sería que el tratamiento A reduzca 2 mmHg más que el fármaco B la tensión arterial sistólica, clínicamente irrelevante.

Cada una de estas etapas de la lectura del ensayo clínico se completa con la adecuada interpretación de los análisis estadísticos. En el siguiente apartado explicamos en forma concisa los empleados habitualmente.

EVALUACIÓN E INTERPRETACIÓN DE LOS RESULTADOS

Los análisis estadísticos que vamos a encontrar en un ensayo clínico, además de los descriptivos, serán aquellos relacionados a la comparación entre las ramas de intervención respecto a los objetivos a analizar. Es decir, que para evento beneficioso o perjudicial asociado al tratamiento de control y de prueba se citarán el *odd ratio*, riesgo relativo, curvas de supervivencia y *hazard ratio*, acompañados del intervalo de confianza.

Recordemos que desde el punto de vista práctico, OR y RR tienen una interpretación similar, se los puede leer como «veces más». Por su parte la interpretación del HR nos demanda un cálculo matemático más, pero nada complicado; sólo debemos restar a 1 el valor del HR y multiplicar el resultado por 100 %. Por ejemplo si el ensayo clínico en el cual se compara espironolactona contra placebo en personas con insuficiencia cardiaca severa nos informa HR=0.70 para mortalidad, hacemos $[(1-0.70) \times 100]$. Entonces el tratamiento con espironolactona disminuye el riesgo relativo de muerte en un 30 % respecto al placebo.

Otros datos importantes que nos van a brindar los estudios clínicos son:

- Reducción de riesgo relativo: consiste en establecer en qué proporción se reduce el riesgo de sufrir el evento en presencia del factor, con respecto al riesgo de la población no expuesta. Para calcularlo se puede realizar $(1-RR) \times 100$ ó mediante la diferencia entre las proporciones de eventos en el grupo control (PEC) y del grupo experimental (PEE) dividido el porcentaje de eventos en el grupo control $(PEC-PEE) / PEC$.
- Reducción de riesgo absoluto (RAR): es la diferencia entre los grupos de RA no expuestos y RA expuestos $(RA_{NE} - RA_E)$ ó entre las proporciones de eventos en el grupo control (PEC) y del grupo experimental (PEE) $\rightarrow PEC-PEE$
- Número Necesario a Tratar (NNT): $1/RAR$. Es un cálculo mediante el cual se determina cuantas personas requieren del tratamiento para evitar el desarrollo del evento.

Ejemplifiquemos estos parámetros mediante la siguiente situación en la cual a personas adultas mayores hipertensas de 5 años de evolución y sin otra comorbilidad se le indicó a un grupo enalapril y a otro placebo durante 7 años a fin de valorar si el fármaco disminuía la tasa de eventos coronarios.

TABLA 20. RESUMEN DE DATOS DE EVENTOS CORONARIOS EN RELACIÓN CON EL USO DE ENALAPRIL

	Pacientes del grupo de estudio (Enalapril) n=100	Pacientes del grupo de control (Placebo) n=130
Síndrome coronario	15	63
Porcentaje	15 %	48,46 %

- $RRR = (PEC - PEE) / PEC \rightarrow (48,46 - 15) / 48,46 \rightarrow 0,69$
El valor de la reducción de riesgo relativo se expresa en porcentaje, por lo cual al resultado se lo debe multiplicar por 100. $0,69 \times 100 \rightarrow 69\%$.
Esto quiere decir que el tratamiento con enalapril redujo el riesgo de presentar síndrome coronario agudo en 69% en comparación con los que recibieron placebo.
- $RRA = PEC - PEE \rightarrow 48,46 - 15 \rightarrow 33,46\%$
El resultado implica que de cada 100 tratados con enalapril, 33,46% no desarrollarán síndrome coronario.
- **NNT.** Para su cálculo el valor de RRA no debe estar expresado en porcentaje. La fórmula es $1/RRA$, entonces vamos a tener $1/0,33 \rightarrow 3,03$.
Básicamente este valor refiere que es necesario tratar con enalapril 3 sujetos con hipertensión arterial durante 7 años para evitar que 1 desarrolle síndrome coronario.

Cuando se valoran los efectos no deseados de la intervención se calcula el incremento del riesgo absoluto (IRA) que se determina de igual forma que el RRA, sin embargo el resultado será un valor negativo. Por ejemplo si en vez de 15% la tasa de eventos coronarios en el grupo con enalapril el valor fuera 71% (mayor al del grupo placebo) tendríamos el siguiente dato:

- $IRA = PEC - PEE \rightarrow 48,46 - 71 \rightarrow -22,74\%$
Esto quiere decir que de cada 100 personas hipertensas tratadas con enalapril, 22,74% tendrán síndrome coronario.

Al ser un valor negativo que nos dice que al recibir tratamiento la persona sufre un daño, en vez de calcularse el NNT, se obtiene el número necesario a dañar (NND), es decir la cantidad de personas que se requieren tratar para que una sufra el efecto no deseado. Al igual que en el caso de NNT, para NND el valor de IRA no debe estar expresado en porcentaje.

- $NND = 1/IRA \rightarrow 1/0,22 \rightarrow 4,54$
Ante esto, por cada 5 personas tratadas con enalapril, 1 va a sufrir un evento coronario.

En conclusión, si bien es una situación ejemplo los datos empleados sobre enalapril y síndrome coronario, los análisis mencionados son importantes a la hora de interpretar y extrapolar los resultados de un ensayo clínico. Deben de tenerse siempre presentes para la valoración del riesgo-beneficio de la intervención y su aplicabilidad.

13 Estudios diagnósticos

Caso clínico

Luego de un mes de que Alberto fuera externado, asiste a la guardia por dolor precordial que se inició hace 30 minutos y cuyas características impresionan de un síndrome coronario agudo. No impresiona ser dolor precordial de otra entidad nosológica. Se le realiza un electrocardiograma intra-angor y determinación de enzimas cardíacas, todo lo cual resulta negativo. Sin embargo, el médico de guardia clínica, considerando el antecedente de un infarto previo, le solicita una angiografía tomográfica computada.

GENERALIDADES

La pregunta que se puede conformar en el marco de la situación problema es cuán útil es la determinación de la angiografía tomográfica computada para realizar el diagnóstico de infarto agudo de miocardio.

En el capítulo anterior describimos cómo se lleva adelante la evaluación de la eficacia de una nueva intervención para un problema de salud. Se podría decir, en forma esquemática, que los estudios de validación de pruebas diagnósticas son similares. Mediante los mismos se busca determinar si un método complementario de diagnóstico clínico es eficaz para una entidad nosológica en comparación con un *gold es standard*, esto es, un método diagnóstico de referencia.

Los estudios de validación de pruebas diagnósticas se caracterizan por ser trabajos científicos analíticos en los cuales el grupo de investigadores efectúa a sujetos sanos y a otros con una cierta enfermedad la prueba diagnóstica establecida para la misma (prueba patrón o *gold standard*), el test diagnóstico en estudio y luego procede a cotejar los resultados. En concreto, son trabajos científicos observacionales, prospectivos y comparativos.

Al momento que una persona nos plantea su problema, con las respectivas manifestaciones sintomáticas, nuestro cerebro de modo automático comienza a hacer un listado de diagnósticos posibles y de quehaceres necesarios para

reducirlo a una sola enfermedad. Es decir, armamos un rompecabezas para determinar la probabilidad de que la persona que nos consulta tenga una determinada patología. Conocer la prevalencia de las enfermedades en el lugar en el cual desempeñamos nuestra labor es muy importante debido a que esta información epidemiológica es un pilar fundamental para desarrollar el planteo de probabilidades diagnósticas. Por ejemplo, si estamos trabajando en un medio en el que la tuberculosis es una enfermedad inexistente y nos consulta un paciente por tos de más de 15 días de evolución y que no estuvo en lugares en los que esta patología es frecuente, pues tuberculosis no sería uno de los principales diagnósticos diferenciales.

Esto se conoce como probabilidad pretest, la posibilidad de que una persona padezca una determinada condición antes de contar con los resultados de los test diagnósticos. Estimar la probabilidad pretest es útil para escoger la prueba complementaria a usar, interpretar de manera adecuada los resultados y decidir la necesidad de iniciar tratamiento (luego de los resultados de las pruebas diagnósticas o de forma empírica).

Como se puede apreciar, de la relación pretest y la eficacia de la prueba diagnóstica empleada surge el diagnóstico definitivo conocido como postest. Por lo tanto, aquí juegan un rol muy importante los estudios de validación de las pruebas diagnósticas ya que nos aportan los datos necesarios para saber cuál método complementario es adecuado usar y en cuánto ayudan a establecer el postest.

PAUTAS PARA LA LECTURA CRÍTICA E INTERPRETACIÓN DE ANÁLISIS ESTADÍSTICOS

En el Capítulo 5 nos hemos referido a los test estadísticos empleados habitualmente para valorar las pruebas diagnósticas. En esta oportunidad vamos a complementar la información brindada oportunamente con los puntos generales que el lector debe de considerar para analizar de modo crítico este tipo de trabajo científico.

Como dijimos, los estudios de validación de pruebas diagnósticas son trabajos comparativos entre la prueba patrón y la de estudio para la determinación de una entidad nosológica particular. Por tal motivo, se van a incluir en el trabajo personas sanas (verdaderos negativos - VN) y enfermas (verdaderos positivos - VP). Si la prueba en estudio arroja un resultado negativo pero la persona padece la enfermedad (prueba patrón positiva) se considera falso negativo (FN); por el contrario, si es positivo y la persona es sana (prueba patrón negativa) es un falso positivo (FP). En la siguiente tabla se esquematiza el cotejo de los resultados de las pruebas complementarias.

TABLA 21. TABLA DE CONTINGENCIA PARA DETERMINACIÓN DE SENSIBILIDAD Y ESPECIFICIDAD DE PRUEBA EN ESTUDIO

	Prueba patrón positiva	Prueba patrón negativa
Prueba en estudio +	VP	FP
Prueba en estudio -	FN	VN

VP: verdadero positivo, FP: falso positivo, VN: verdadero negativo, FN: falso negativo

Como es un estudio comparativo y la eficacia de la nueva prueba surge de su cotejo con la prueba patrón, el primer interrogante que debemos hacernos al leer un estudio de validación es si la prueba *gold standard* es la conveniente. Si se emplea una prueba patrón que no es la óptima para el diagnóstico de la enfermedad, no vamos a poder confirmar que la prueba en estudio esté midiendo con la exactitud adecuada lo que debe valorar.

Luego debemos corroborar que la muestra de sujetos incluidos en el estudio sea la correspondiente para el objetivo planteado en el estudio. Recordemos que el pretest de una enfermedad está sujeto a la prevalencia de la misma. Por lo tanto, para una correcta evaluación de la calidad diagnóstica de la prueba complementaria es recomendable que la proporción de sujetos con la condición de interés (enfermedad) sea similar a la prevalencia de la patología. La comparación entre las pruebas se hace en un grupo de personas con la entidad nosológica que se quiere diagnosticar y otro grupo con sujetos sanos. Este último no implica la ausencia de enfermedad. En muchas ocasiones, la validación de la prueba diagnóstica se hace con el objetivo de valorar la eficacia del test diagnóstico para discernir diferentes grados de la misma enfermedad o cuadros patológicos de presentaciones similares.

Una vez que analicemos el objetivo del estudio, las pruebas a emplear y la muestra de sujetos incluidos, debemos valorar las características operativas de la prueba complementaria en estudio, es decir, el resultado de la comparación con el test de patrón. Los parámetros incluyen:

- Sensibilidad: es la capacidad de la prueba para identificar a las personas que tienen la condición de interés (enfermos). Se determina mediante la siguiente fórmula $VP/VP+FN$ y se expresa en porcentaje. Si tenemos 100 sujetos enfermos y la sensibilidad del test es de 71 %, quiere decir que detecta correctamente a 71 personas con la patología y 29 personas enfermas son incorrectamente calificadas como sanas (falsos negativos).
- Especificidad: de manera opuesta al punto anterior, especificidad es la capacidad de clasificar en forma debida a las personas sanas (sin la condición de interés). Se determina mediante la siguiente relación: $VN/FP+VN$ y también se expresa en porcentaje. Si la prueba tiene una especificidad del 85 %, quiere decir que cada 100 personas sanas 85 tendrán un resultado negativo del test y 15 positivo (falsos positivos).

La determinación de sensibilidad y especificidad detallada en los párrafos anteriores se enfoca en una modalidad dicotómica de la prueba diagnóstica, entiéndase positivo o negativo, y queda de lado la gama de estudios complementarios que se miden en escala continua (o discreta ordinal). Existe, pues, un test estadístico que permite valorar, ante esta situación, el o los puntos de corte de la prueba diagnóstica complementaria. Se denomina análisis de curvas ROC (*receiver operating characteristic*). Se trata de gráficas en las que se representa la sensibilidad de una prueba en función de los falsos positivos para distintos puntos de corte, por lo tanto nos brindan los valores de sensibilidad y especificidad.

Pasemos a ejemplificar una curva ROC. Para tal fin emplearemos el nivel del anticuerpo contra el receptor β_1 adrenérgico en personas con enfermedad de Chagas para el diagnóstico de insuficiencia cardíaca. La prueba patrón fue la ecocardiografía.

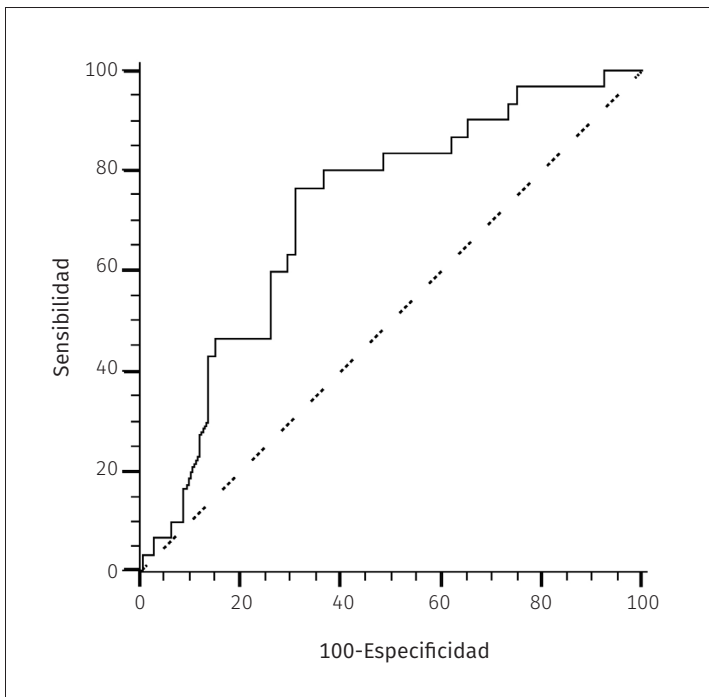


FIGURA 14. CURVA DE SENSIBILIDAD Y ESPECIFICIDAD

El valor óptimo de nivel de anticuerpo para el diagnóstico de insuficiencia cardíaca es de 1,18, con una sensibilidad de 76,7 % (IC 95 % 57 %-90 %) y una especificidad de 69 % (IC 95 % 60 %-77 %). Como se puede ver en la gráfica de la curva ROC, la misma tiene un comportamiento creciente, propiedad que demuestra la relación inversa entre sensibilidad y especificidad. Si la intención de una prueba diagnóstica es identificar a la mayoría de las personas con la condición de interés, se puede modificar el punto de corte de la prueba incrementando la sensibilidad. Esto implicaría que el valor de especificidad disminuya.

Otro dato que nos brinda el análisis de la curva ROC es el área debajo de la curva (AUC). En nuestro caso, esta fue de 0,7 (IC 95 % 0,6-0,8; $p < 0,0001$). Su valor nos permite conocer la calidad del test para el propósito evaluado. El rango de la misma es de 0,5 a 1, en tanto 1 es el valor diagnóstico perfecto y 0,5 implica que la prueba no tiene capacidad discriminatoria. Sobre la base de este intervalo se conforman 5 grupos de clasificación de pruebas diagnósticas:

- 0,5 a 0,6: mala capacidad discriminatoria diagnóstica.
- 0,61 a 0,75: el test diagnóstico presenta regular calidad.
- 0,76 a 0,9: la eficacia discriminatoria de la prueba diagnóstica es buena.
- 0,91 a 0,97: el test es muy bueno.
- 0,98 a 1: la prueba diagnóstica presenta excelente capacidad discriminatoria.

Por otra parte, en el ejercicio de nuestra profesión, sensibilidad y especificidad no son los determinantes a considerar para saber si el resultado de un estudio complementario que le hemos indicado a una persona es verdadero o falso. Influyen en este sentido el contexto clínico y los valores predictivos del estudio.

El valor predictivo se puede definir como la probabilidad de que el resultado de la prueba complementaria diagnóstica se corresponda en verdad con la situación del paciente, es decir:

- Valor predictivo positivo: nos dice que probabilidad real tiene el sujeto de estar enfermo si la prueba de estudio le dio positiva $\rightarrow VP/VP+FP$.
- Valor predictivo negativo: si el resultado de la prueba en estudio fue negativo, mediante la determinación del valor predictivo negativo se calcula la probabilidad de que el sujeto realmente no esté enfermo $\rightarrow VN/FN+VN$.

Los valores de sensibilidad y especificidad son condiciones intrínsecas de la prueba diagnóstica y definen su eficacia en forma independiente de la prevalencia de la condición de interés. Sin embargo, al no estar relacionados con la prevalencia de la enfermedad (y por lo tanto influenciados por el pretest) no intervienen de manera categórica a la hora de tomar una decisión sustentada en resultado de la prueba diagnóstica.

Ahora bien, recién mencionamos al valor predictivo (positivo o negativo) como herramienta que nos posibilita verificar que el resultado de la prueba

diagnóstica sea el «correcto» en virtud del contexto clínico, lo que juega un rol importante en el accionar profesional. Pero presenta una limitante muy importante. Su valor está muy ligado a la proporción de los sujetos con la condición de interés en el estudio de la validación de la prueba diagnóstica, por lo tanto es dependiente de la prevalencia de la enfermedad.

Pasemos a ejemplificar esta situación suponiendo que tenemos un test cuya sensibilidad es de 75,7 %, su especificidad del 97 %, y la prevalencia de la enfermedad es 82,5 casos por cada 1000 habitantes.

TABLA 22. EJEMPLIFICACIÓN DE PRUEBA DIAGNÓSTICA, PRIMERA PARTE

	Prueba patrón +	Prueba patrón -	Total
Prueba en estudio +	50	22	72
Prueba en estudio -	16	712	728
Total	66	734	800

El valor predictivo positivo es del 69 % y el negativo del 97 %. En este contexto, si el resultado del test es positivo, la probabilidad de que la persona presente la enfermedad es de moderada a alta.

Ahora modifiquemos la prevalencia a 8,25 casos por cada 1000 habitantes, con lo que nuestra tabla quedaría de la siguiente forma:

TABLA 23. EJEMPLIFICACIÓN DE PRUEBA DIAGNÓSTICA, SEGUNDA PARTE

	Prueba patrón +	Prueba patrón -	Total
Prueba en estudio +	5	24	29
Prueba en estudio -	2	769	771
Total	7	793	800

El valor predictivo negativo es del 99 %, mientras que el valor predictivo positivo disminuyó a 17 %. En este escenario, si el resultado del test es positivo, no estamos en condición de aseverar que realmente lo sea.

Puesto que la prevalencia influye en la determinación e interpretación de los valores predictivos, no pueden extrapolarse del trabajo científico de valoración de estudios diagnósticos a la práctica clínica. Por tal motivo, se requiere calcular la razón de verosimilitud o razón de probabilidad (*likelihood ratio*).

- *Likelihood ratio* (LR) es el cociente de probabilidad de obtener un resultado mediante la prueba en estudio, esto es, cuánto se modifica el pretest, y puede ser LR+ en caso de presentar un test positivo o LR- si el resultado es negativo. Se calcula de la siguiente forma:

LR+: sensibilidad/(1-especificidad)

LR-: (1-sensibilidad)/especificidad

Para interpretar cómo modifica el LR nuestro pretest se puede emplear el Nomograma de Fagan. El mismo consta de 3 columnas; la primera es la probabilidad pretest del diagnóstico de la condición de interés, la del medio corresponde al valor de LR (sea positivo o negativo) y la última representa el valor diagnóstico postest.

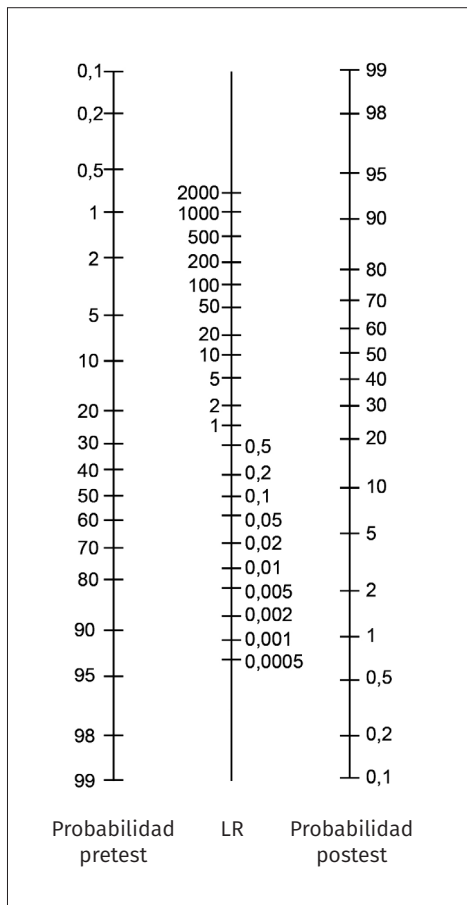


FIGURA 15. NOMOGRAMA DE FAGAN

Otra forma más sencilla de interpretarlo, de no contar con el Nomograma, es por ejemplo: si el LR+ del estudio es de 4 para la condición de interés, el paciente con dicho resultado tiene 4 veces más probabilidades que una persona sana de tener la enfermedad. El intervalo de confianza del LR debe evaluarse respecto de su amplitud. Si el mismo es estrecho, implica que no existió mucha dispersión de los resultados obtenidos con la prueba en estudio en comparación con el *gold standar*.

El último punto que debe responder un profesional de la salud ante un artículo de validación de pruebas diagnósticas es el interrogante sobre cuál sería la utilidad de su aplicación en cuanto a la relación costo-beneficio en el medio en que trabajamos.

RESOLUCIÓN DE LA SITUACIÓN CLÍNICA

En esta oportunidad, el caso clínico disparador del tema tratado a lo largo del capítulo es sobre la validez del *score* de calcio coronario como método diagnóstico de infarto agudo de miocardio. Si usamos las palabras clave «acute chest pain», «cardiac CT», uno de los artículos hallados es «Coronary Computed Tomography Angiography For Early Triage of Patients with Acute Chest Pain – The Rule Out Myocardial Infarction Using Computer Assisted Tomography (ROMICAT) Trial», publicado por UDO Hoffmann y colaboradores en el año 2009 en el *Journal of the American College of Cardiology*.

El trabajo de investigación tenía como fin determinar la utilidad de la angiografía tomográfica computada como método complementario para determinar si una persona que consulta por dolor de pecho en un servicio de guardia, con electrocardiograma y valor de enzimas cardíacas normales, tiene o no un síndrome coronario. La sensibilidad y el valor predictivo negativo para síndrome coronario agudo, en el contexto de estenosis significativa de las arterias coronarias, fue del 77 % y 98 % respectivamente; por su parte, la especificidad fue de 87 %.

Con estos valores citados podemos establecer la razón de probabilidad positiva (LR+) para extrapolar los resultados a nuestra situación problema. El valor de LR+ es de 5,9. Apliquemos este valor en el Nomograma de Fagan considerando una probabilidad pretest moderada ya que no hay otro diagnóstico presuntivo, pero inicialmente tenemos un electrocardiograma intrangor negativo.

Como se observa, si partimos de una probabilidad pretest del 50 %, para un LR+ de 5,9, la probabilidad postest de que Alberto tenga un síndrome coronario agudo, de resultar positiva la angiografía tomográfica computada, es superior al 80 %.

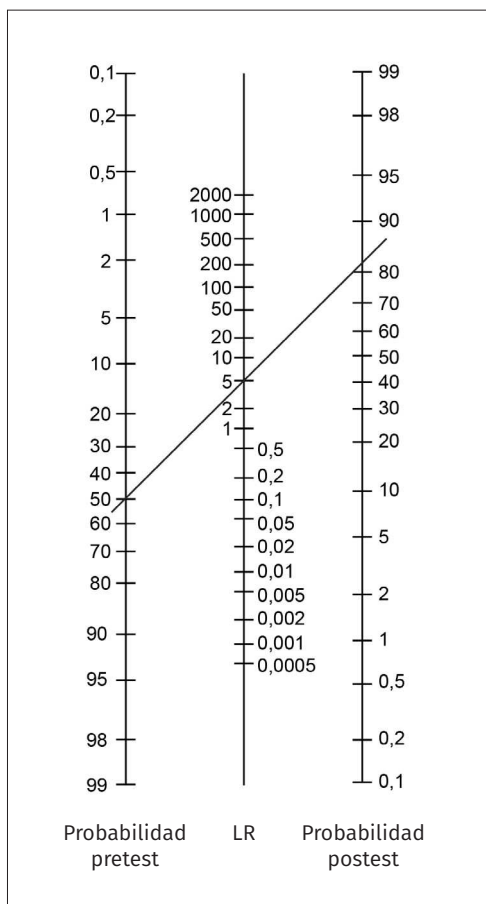


FIGURA 16. NOMOGRAMA DE FAGAN APLICADO EN LA RESOLUCIÓN DE CASO CLÍNICO

14 Metaanálisis

Caso clínico

Si bien el médico de Alberto ha encontrado antecedentes bibliográficos que prueban la relación entre el nivel de ácido úrico y eventos cardiovasculares, en la última jornada de medicina asistió a una charla en la cual un grupo de investigadores prestigiosos comentaba recientes publicaciones en las cuales no se apreciaba relación entre ácido úrico y el desarrollo de patologías cardiovasculares.

GENERALIDADES

Al ser el arte médico una disciplina científica, queda sujeto a las características que son propias del saber científico. En tanto, destacamos que el conocimiento debe estar sujeto al escepticismo, es decir, a ser juzgado y revalorado en forma continua, libre de ser ratificado o rectificado. Esto lleva a que en forma simultánea e independiente unos de otros, diversos investigadores trabajen sobre un mismo tópico. Por lo tanto, es muy probable que, además de encontrarnos con mucha información, ésta presente observaciones discordantes. De ser así, compararemos la metodología aplicada en virtud de nuestra pregunta clínica, pero podemos volver al punto de partida ya que trabajos científicos de calidad que nos muestran resultados controversiales.

Ante este panorama oscuro dentro de la MBE, la luz al final del túnel surge de las revisiones sistémicas cuyo objetivo es resumir e integrar la evidencia respecto de una pregunta específica. Existen dos tipos de revisiones sistémicas. Por un lado, nos encontramos con la revisión cualitativa, en la cual se sintetizan las diferentes observaciones con los pro y contras de los estudios para que el lector, sobre la base de la situación clínica particular que demandó la búsqueda, realice su juicio crítico de aplicabilidad. Por otro lado, está la revisión es la cuantitativa o metaanálisis, que supone el uso de análisis estadísticos con combinación de los resultados de diversos estudios en un solo estimador. Este tipo de revisión es el que vamos a desarrollar en este capítulo.

DEFINICIÓN, VENTAJAS Y LIMITACIONES

El metaanálisis es una síntesis analítica de diversos trabajos científicos realizados con el propósito de contrastar la misma hipótesis (o muy similares) y es el nivel de evidencia más alto. Al hacer un *pool* de trabajos científicos se incrementa la potencia del estudio y es factible determinar con más exactitud la existencia real de diferencia entre un fármaco y otro para el tratamiento de una enfermedad, la asociación de un factor de riesgo para con el desarrollo de un problema de salud, o la eficacia de un método diagnóstico complementario. Además, reduce los sesgos que pudiera presentar en forma independiente cada trabajo incluido en la revisión y mejora así la validez interna, de modo tal que sea posible extrapolar sus resultados de manera adecuada para responder la pregunta clínica.

El objetivo y la población objeto de estudio del metaanálisis enmarcan el tipo de trabajo científico a incluir y la metodología de búsqueda que se emplea para recabar las publicaciones. Se pueden utilizar los resultados publicados de los estudios que se incluyan o los datos individuales. Esto último implica analizar la base de datos original y, si bien ofrece una mayor calidad de resultados al permitir ampliar la exploración estadística de las observaciones recolectadas, es necesario contar con la colaboración de los diversos grupos de investigadores de los diferentes estudios.

No obstante del método a emplear, es importante recordar que existen dos grandes bases de datos científicas, *Medline* y *Embase*, que comparten aproximadamente un 34 % de publicaciones, por lo cual si el grupo de investigadores se limita a una sola base está dejando fuera del análisis una cantidad importante de artículos científicos que pueden influir en el resultado final de la revisión. Como se advierte, la calidad de la revisión se encuentra sujeta al proceso de búsqueda, a la selección de los artículos y al análisis de los mismos.

Por otra parte, aunque el metaanálisis es considerado el mayor nivel de evidencia, tampoco está exento de controversias. En consideración del proceso de desarrollo de la revisión, es factible que dos metaanálisis sobre el mismo tema y con igual objetivo presenten conclusiones dispares debido a que tienen criterios de inclusión diferentes. También puede suceder que, con posterioridad a la comunicación de sus resultados, surja, por ejemplo, un nuevo ensayo clínico hecho aleatorizado de mayor tamaño muestral y buena calidad metodológica que contradiga las observaciones previas. En estos casos hay tendencia a considerar las conclusiones del nuevo estudio, sin embargo, es el profesional quien debe de analizar la validez externa de ambos estudios con relación a la situación clínica que gatilló la pregunta.

Otra limitación importante del metaanálisis se vincula en forma directa con el sesgo de publicación. Este se refiere a que las revistas científicas suelen

dar lugar a artículos que presentan resultados positivos y este contexto favorece que los propios resultados de la revisión se encuentren sesgados, con tendencia a rechazar la hipótesis nula.

El motivo ya lo había enunciado Francis Bacon en el siglo XVII al expresar en su *Novum Organum*: «El intelecto humano se conmueve más por las afirmaciones que por las negaciones». En cuanto a esto, por parte de los investigadores se reconoce la falta de entusiasmo ante resultados que pueden no ser los esperados, la presión de publicar asociaciones importantes y positivas, así como el hecho de que los hallazgos negativos irán a comunicarse seguramente a revistas con menor factor de impacto. En la revisión por pares y editores de revistas, es frecuente que se seleccionen solo los resultados positivos ya que atraen más lectores.

Pero, ¿qué problema puede traer el sesgo de publicación? Puede producir estimaciones erróneas de la magnitud del efecto de las intervenciones estudiadas en los metaanálisis (por ejemplo, sobreestimar beneficios o subestimar efectos adversos). No olvidemos que esta tendencia afecta a todas las ramas del conocimiento biomédico, máxime sabiendo que aproximadamente la mitad de los ensayos clínicos randomizados controlados no se publican por este motivo. Sin duda, esta ausencia de información puede conducir a consecuencias inesperadas en las decisiones que los que estamos del otro lado tomamos a partir de los resultados de los metaanálisis, por más prolijos que puedan resultar en otros aspectos. Por eso se ha dado en llamar «el talón de Aquiles» de los metaanálisis. Cabe mencionar que, como estrategia ante la sospecha de sesgo de publicación, se ha propuesto un modo de búsqueda conocido como «Grey literature search» (o búsqueda de literatura gris), que hace referencia a recurso para encontrar el cuerpo de información que no se halla circulando a «simple búsqueda», como contactar a investigadores originales, ingresar a disertaciones, trabajos de tesis o resultados preliminares sobre el tema, revisar reportes técnicos o los registros de ensayos clínicos que son públicos. De esta manera, uno puede constatar si en verdad hubo otros trabajos vinculados que fueron excluidos debido a sus resultados.

En conclusión, si bien las revisiones sistémicas, particularmente las de diseño cualitativo (metaanálisis) se posicionan como la «última» palabra de la medicina basada en evidencia, no están exentas de posibles falencias metodológicas o sesgos que hacen sumamente necesario a la lectura crítica e interpretativa por parte del profesional de salud.

PAUTAS PARA LEER UN METAANÁLISIS

Los interrogantes que el lector se debe plantear ante una revisión sistémica, en particular un metaanálisis, para evaluar su validez se pueden resumir en los siguientes puntos:

- ¿Se limita específicamente una pregunta clínica? Dijimos que una revisión era el análisis integrado de artículos científicos sobre un tema puntual. Por lo tanto, el investigador que desarrolle este tipo de estudio debe de hacerse una pregunta clínica determinada que enmarca al objetivo principal y da inicio al proceso de búsqueda científica, selección de artículos, y lectura crítica de los mismos.
- ¿Cuáles son los criterios de inclusión y exclusión de los estudios? De igual forma que en otros trabajos científicos, se deben delimitar las características de la muestra poblacional (artículos científicos) que va a participar del estudio. Así, los autores de la revisión tienen que detallar las características de los artículos que contextualicen a los sujetos, tipo de observaciones/ intervenciones, comparaciones y resultados que darán respuesta a la pregunta clínica. Si este punto no tiene coherencia con el anterior es probable que la revisión presente sesgos que juegan en contra de las observaciones del estudio.
- ¿La metodología de la búsqueda científica fue apropiada en cuanto al objetivo del estudio? Para continuar con el proceso de la inclusión de los artículos científicos y responder la pregunta clínica de la revisión se debe de especificar la metodología de la búsqueda científica con inclusión de las palabras clave y las bases de datos empleadas. Considerando que se está realizando una revisión sistémica que integra varias publicaciones, es preferible que no existan límites de búsqueda que puedan promover la exclusión indebida de artículos.
- ¿Los artículos encontrados, además de cumplir con los criterios de inclusión, poseen calidad metodológica y validez interna? Obtenidos los resultados de la búsqueda bibliográfica comienza la etapa de tamizaje de los artículos. Como se emplea más de una base de datos, es probable que tengamos resultados por duplicado, por lo cual se realiza un primer *screening* con la finalidad de eliminar aquellos trabajos repetidos. En segunda instancia, los investigadores analizan tanto título como resumen de los trabajos con el propósito de incluir aquellos relevantes al objetivo de la revisión. Finalmente, se concreta el análisis del texto completo (*full text screening*) y se valora la coherencia interna de los trabajos en relación con el objetivo de los mismos. Al final del metaanálisis, por lo general se colocan en tablas anexas los datos más relevantes de cada estudio incluido para que el lector pueda hacerse una idea de su metodología, tamaño muestral y magnitud de efecto.
- ¿Existe sesgo de publicación e inclusión de los artículos? Para analizar este punto, en todo metaanálisis debería constar una gráfica denominada *funnel plot* (de pirámide o embudo invertido) en la cual cada punto representa el trabajo científico incluido. Este tipo de análisis fue propuesto por Egger en el año 1997 como una manera de poner en evidencia de manera simple la calidad de las revisiones. En uno de sus estudios encontró que un 38 % de

los metaanálisis de revistas de punta presentaban este sesgo, al tiempo que solo halló un 13 % en las publicaciones de la iniciativa Cochrane. En el eje X se detalla la magnitud del efecto valorado y en el Y el tamaño muestral del estudio. A continuación ejemplificamos dos gráficas de *funnel plot*.

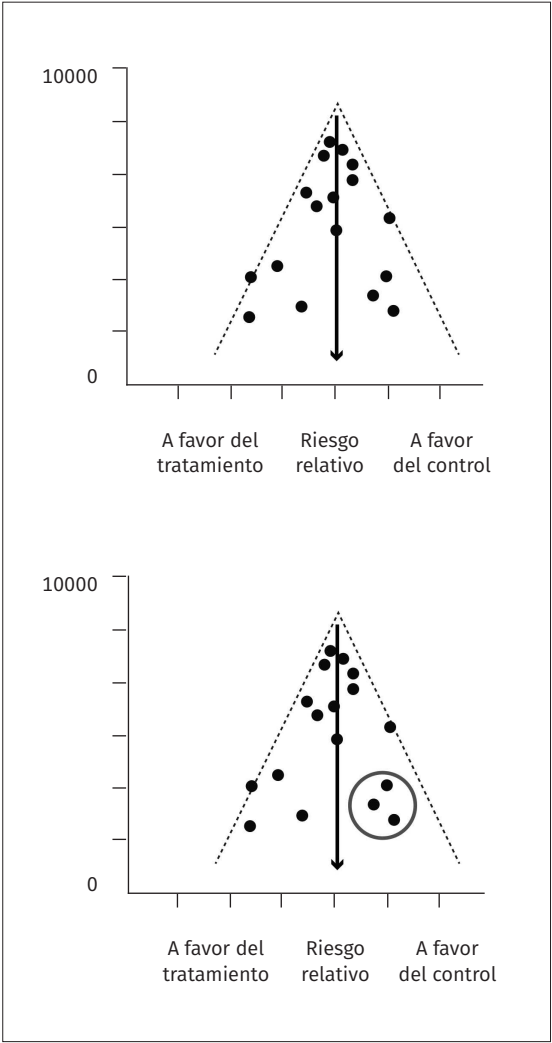


FIGURA 17. EJEMPLO DE GRÁFICA FUNNEL PLOT

Cuando el embudo es asimétrico, es decir con mayor cantidad de puntos de un lado que del otro como en la segunda situación, el lector debe considerar la posibilidad de que exista sesgo de publicación y/o de inclusión. También hay otras técnicas cuantitativas más complejas, como el test de Egger, de correlación lineal, pero desde el punto de vista práctico, el *funnel plot* es lo más utilizado.

- ¿Los resultados de los estudios incluidos son factibles de ser combinados? Para poder integrar los datos de los trabajos científicos en la revisión no debe haber grandes diferencias entre ellos. Para esto tenemos pruebas estadísticas de valoración de heterogeneidad. Una de las más empleadas es el test de Cochran, basado en χ^2 considerado significativo cuando el valor de p es menor o igual a 0,1; si esto sucede indica que hay diferencias entre los resultados de los estudios incluidos.
- Otra prueba estadística para valorar este punto es I^2 , que mide la proporción de la variación. Su resultado se expresa en porcentaje y, a mayor valor, mayor heterogeneidad; valores de 25 %, 50 % y 75 % refieren a heterogeneidad baja, moderada o muy alta. Por lo general, cuando se supera el 50 % es preferible no hacer la combinación de los resultados para determinar un valor del efecto resumen ya que los estudios poseen demasiadas variaciones entre sí.
- ¿Cuál es el resultado de la revisión? Quien lee una revisión debe de conocer cuál es la magnitud del efecto estudiado. Es habitual que esto se publique en una gráfica denominada *forest plot*, en la que se detallan el OR, RR, HR o NNT, dependiendo de lo que se evalúe, en forma independiente para cada trabajo incluido y del muestreo global (combinación de los trabajos científicos), siempre acompañado del respectivo intervalo de confianza. A continuación presentamos un esquema de *forest plot* para facilitar el reconocimiento de sus elementos más relevantes a la hora de interpretarlos. Imaginemos que se trata de 4 trabajos revisados que estudian el efecto de un tratamiento sobre un grupo de pacientes experimental en comparación con un grupo control.

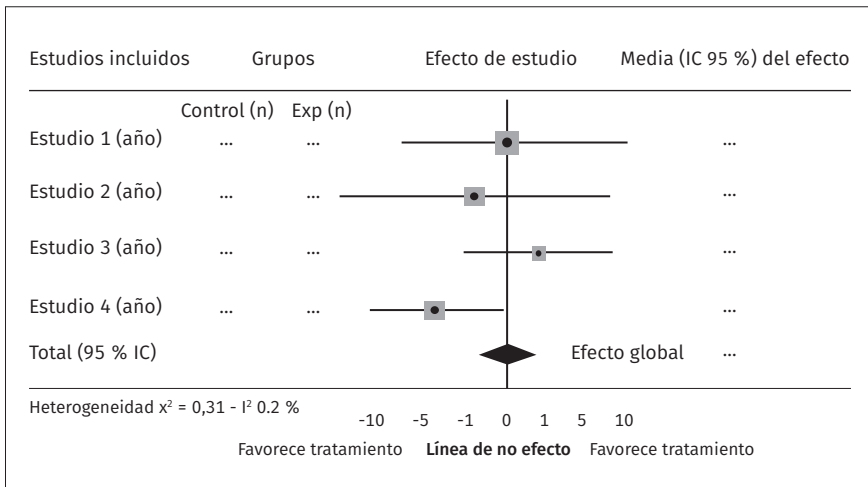


FIGURA 18. EJEMPLO DE GRÁFICA FOREST PLOT

Se observa que a cada trabajo le corresponde un trazo que atraviesa la línea de no-efecto (por ejemplo, $RR=0$). Dicho trazo tiene un punto central que indica el valor medio del parámetro de magnitud de efecto utilizado, mientras que la línea hacia ambos lados expresa su intervalo de confianza al 95 %. A su vez, el tamaño de la caja central (en naranja) representa el «peso» total del estudio en el resultado final atento a su tamaño muestral. El último renglón indica el estadístico global, que es una sumatoria de los efectos. En este ejemplo, vemos que la intervención se posicionaría en 0, por lo cual no estaría revelando efecto alguno sobre la patología de estudio. Es importante analizar estos gráficos ya que permiten visualizar cada trabajo incluido y su resultado en pocos minutos. Además, es conveniente no olvidar los intervalos de confianza de cada efecto ya que, por ejemplo, en este caso, si vemos el trabajo N° 3, que parece discordar con el resto, observamos que su efecto se posiciona del lado que indicaría que es preferible no tratar tal vez por efectos adversos o falta de reducción del *end point* ($RR >1$). Sin embargo, su intervalo de confianza sobrepasa al 0 también en sentido opuesto, lo que da idea de que en algunos pacientes el tratamiento sí aportó beneficios.

Como en el resto de los tipos de trabajos científicos, la lectura crítica de la revisión concluye con la evaluación de la validez externa, es decir, si las observaciones del metaanálisis son aplicables a la situación problema que motivó la búsqueda científica. De no ser así, se debe de revisar el proceso de búsqueda.

Cabe mencionar que existen entidades que se dedican al desarrollo de revisiones sistémicas o metaanálisis en forma independiente de otras entidades (libre de conflictos de interés), por ejemplo, la Colaboración Cochrane (<http://www.cochrane.org>), en cuya biblioteca podrá encontrar revisiones de diversos temas de la práctica habitual clínica.

RESOLUCIÓN DE LA SITUACIÓN PROBLEMA

Como se advierte, la situación problema al inicio de este capítulo muestra la controversia con la que cualquier profesional se puede encontrar con relación a diversos aspectos. Cuando usamos las palabras clave «middle-aged men», «uric acid level», «risk factor for cardiovascular disease», encontramos el artículo denominado «Uric Acid Level as a Risk Factor for Cardiovascular and All-Cause Mortality in Middle-aged Men», publicado en *Archives of Internal Medicine* en el año 2004. Se trata de un trabajo cohorte en el cual se observa que el valor sérico del ácido úrico está ligado al desarrollo de muerte cardiovascular.

No obstante, estos resultados no están ajenos a observaciones dispares. De Vera y colaboradores, mediante un estudio de cohorte poblacional denominado «Independent impact of gout on the risk of acute myocardial infarction among elderly women: a population-based study», describieron que en las mujeres el ácido úrico sérico no se asocia al desarrollo de infarto fatal (RR 1,33, IC 95 % 0,99 a 1,78), y en el caso de los hombres directamente no se relaciona al padecimiento de infarto.

Como este ejemplo, hay otros que evidencian la discordancia con la cual uno se puede encontrar al tratar de dar respuesta a una pregunta clínica. Los dos son trabajos cohortes, metodológicamente correctos, que observan asociaciones distintas respecto del nivel de ácido úrico sérico y el desarrollo de infarto. Entonces, ¿cuál es la verdad?

Nuestro proceder debe ser la reestructuración de la búsqueda científica u orientarnos a niveles de evidencias que subsanen la controversia. Este es el caso de la revisión sistémica. Si limitamos la búsqueda a este tipo de estudio nos vamos a encontrar con el artículo denominado «Gout and Risk of Myocardial Infarction: A Systematic Review and Meta-Analysis of Cohort Studies», publicado en la revista *Plos One* en 2015. El objetivo del estudio es analizar si el nivel de ácido úrico y/o el padecimiento de gota se relacionan con el desarrollo de infarto fatal y no fatal en la población general y por sexo. La búsqueda bibliográfica fue realizada tanto en la base de *Medline* como en la *Embase*, lo cual es bueno porque disminuye el sesgo de inclusión de trabajos. Pero, por otro lado, en la figura 2 de dicho estudio, cuando se

grafican los estudios incluidos, se aprecia que solo uno de ellos presenta resultados negativos, y esto puede estar vinculado al sesgo de publicación.

En la misma figura se detalla también el análisis de heterogeneidad (I^2) cuyo valor es superior al 75 %, punto de corte que refiere que la heterogeneidad de los trabajos es muy alta y debe de alarmarnos en cuanto a que existen disparidades en los estudios incluidos, por lo tanto, la medida resumen del efecto debe ser considerada con cautela y los autores del trabajo de revisión tienen que explicar las razones por las cuales se produce la heterogeneidad.

Al realizarse el análisis en forma estratificada por evento, esto es, desarrollo de infarto fatal y no fatal, en el primer caso el análisis de heterogeneidad muestra que existe homogeneidad entre los estudios y el valor resumen es RR 1,11 (IC 95 % 0,96 a 1,28), lo cual implica que no hay asociación entre gota y el desarrollo de infarto agudo de miocardio fatal. Por su parte, en el caso del desarrollo de infarto no fatal, la heterogeneidad es moderada, levemente menor al 50 %, y por ello es factible la combinación de datos. En tanto, a diferencia del evento anterior, sí se aprecia relación entre gota y el evento en cuestión (RR 1,29, IC 95 % 1,19 a 1,39).

Aunque ya tenemos una parte resuelta, este estudio refiere a la población general con independencia del sexo y no responde a la situación particular de Héctor. Hace un análisis por subgrupo de sexo observando tanto en hombres como en mujeres que la gota se asocia al desarrollo de infarto; sin embargo, hay dos puntos a considerar. En primer lugar, el índice de heterogeneidad en ambos casos es mayor al 75 % y, en segundo lugar, no se analiza el evento en forma estratificada (infarto agudo de miocardio fatal y no fatal).

En conclusión, el trabajo nos ayuda a contestar la situación problema e incluso a saber que, sin importar el sexo, el padecimiento de gota se asocia al desarrollo de infarto. Pero esta observación debe considerarse con las limitaciones que presenta el estudio, que se traducen en la gráfica de embudo en la cual se aprecia la falta de trabajos con resultados negativos y el alto índice de heterogeneidad. Estas dos situaciones resaltan posibles sesgos capaces de influir en el resultado del trabajo.

15 Falacias en medicina

TANGAS, OJOTAS, DORITOS, YOGA Y LENTES DE CONTACTO (ENTRE OTRAS COSAS)

En un centro obstétrico se realizó una encuesta a mujeres embarazadas y posparto sobre 41 preferencias en 6 categorías: comidas, vestimenta, actividades, salud y cosmética, género del hijo y un grupo de misceláneas. Las respuestas debían ser de elección simple: por ejemplo, si usaban anteojos o lentes de contacto, zapatillas de tenis u ojotas, si preferían los autos comunes o las 4 x 4, papas fritas o Doritos, bifés o hamburguesas, manteca o margarina, pollo o carne, Coca Cola o Pepsi, tanga o bombacha común, la playa o la montaña, trotar o practicar yoga, leer libros o mirar películas, el otoño o la primavera, enviar mensajes de texto o de voz, si deseaban tener un niño o una niña, usar tinta azul o negra, joyas de oro o de plata, etcétera.

El embarazo no deseado fue más prevalente entre las mujeres que preferían las tangas, enviar mensajes de texto, comer Doritos, usar lentes de contacto, ver películas, la playa y el yoga. Tuvieron mayor cantidad de niñas las que prefirieron los autos comunes, el pollo y el yoga. El mayor número de niños se dio en las mujeres que preferían las 4 x 4, la carne y deseaban tener niños. Las cesáreas fueron más prevalentes entre las que prefirieron los lentes de contacto, la manteca, la tinta azul y las ojotas. Los valores de p fueron todos menores a 0,05 y en ocasiones de 0,0001.

Si bien parece una broma, el trabajo científico es real y fue publicado por Ruchi P. Moore y colaboradores en el año 2013 bajo el nombre «Thongs, Flip-Flops, and Unintended Pregnancy: The Seduction of $P < 0.05$ ».

¿Cómo es posible que estas asociaciones ocurran? Es obvio que aunque sean estadísticamente significativas son absurdas. Como hemos visto a lo largo del libro, existen distintas falacias que ocurren en la investigación médica y pueden engañar a quien lee el artículo científico.

¿ASOCIACIÓN CASUAL O CAUSAL?

Para establecer la relación de causalidad de una asociación son de utilidad los criterios que propuso en 1965 el eminente epidemiólogo Austin Bradford Hill y que muestran una similitud manifiesta con lo postulado en 1739 por el filósofo inglés David Hume.

- Fuerza de la asociación: mientras mayor sea la relación entre dos variables, mayor será la probabilidad de que exista realmente una relación. Por ejemplo, la tasa de cáncer de pulmón en fumadores es 10 veces mayor que en no fumadores.
- Consistencia: sucede cuando la asociación se observa en distintas circunstancias, personas y tiempo.
- Especificidad: es cuando una misma causa produce siempre el mismo efecto, el que no aparece sin esa causalidad.
- Relación temporal: la causa debe preceder al efecto.
- Curva dosis-respuesta: implica que a mayor exposición al factor causal se incrementa el riesgo de desarrollo de enfermedad. Por ejemplo, a mayor cantidad de cigarrillos fumados, mayor el riesgo de cáncer de pulmón. Puede llegar a ser 20–30 veces mayor que en no fumadores.
- Plausibilidad biológica: tener en cuenta que la plausibilidad depende de los conocimientos sobre un tema determinado, que puede variar con el tiempo. Esto quiere decir que la asociación encontrada no debería contradecir el cuerpo de conocimientos vigente. Sin embargo, es importante aclarar que este criterio no tiene que ser excluyente, ya que la asociación observada puede ser nueva para la ciencia y no debería desecharse sin más. Citan a Sherlock Holmes, quien dirigiéndose a Watson dice: «cuando usted ha eliminado lo imposible, lo que queda, aunque improbable, debe ser la verdad» (Conan Doyle, 2014).
- Coherencia con la historia natural y epidemiología de la afección en cuestión.
- Evidencia experimental: es el nivel de evidencia más sólido para establecer causalidad ya que reúne todos los requisitos necesarios a fin de reproducir la situación ideal en la que un factor desarrolla el efecto. Por cuestiones éticas debe de aplicarse en seres humanos solamente cuando se trata de efectos preventivos. Por ejemplo, el descenso de la PA en la hipertensión reduce la mortalidad.
- Analogía: refiere a que, ante asociaciones causales similares, existe la probabilidad de que aquellas induzcan enfermedades similares. Por ejemplo, si un virus puede producir embriopatía, otro virus también podría hacerlo.

Estos postulados no son suficientes ni necesarios (salvo el de temporalidad) para establecer una relación causal, que debe ser analizada teniendo en cuenta la naturaleza de la asociación en estudio.

Siempre es posible que la relación entre variables sea resultado del azar, y la estadística no puede afirmar la causalidad o casualidad. Sin embargo, sí es capaz de evaluar la probabilidad de que efectos de una determinada magnitud ocurran por azar.

COMPARACIONES MÚLTIPLES: LA INFLACIÓN NO AFECTA SOLO A LOS PRECIOS

Al usar una prueba estadística el valor de p considerado significativo se suele fijar en 0,05, pero al efectuar múltiples pruebas puede aumentar la cantidad de «falsos positivos» (errores tipo I). Supongamos que comparamos las medias de dos muestras de una variable que se distribuya de forma normal. Al hacer la prueba de t de Student, el estadístico informa « $p < 0,05$ », por lo cual se asume que las medias son diferentes, ya que la probabilidad de un resultado significativo es de 0,05 y de un no significativo de 0,95. Pero si se realizan 20 comparaciones independientes, la probabilidad de que ninguna sea significativa será de $0,95^{20} = 0,36$, mientras que la probabilidad de que alguna sí lo sea es de $1 - (\text{Probabilidad de no ser significativo})^n$, lo que es lo mismo, $1 - (1 - 0,05)^{20}$, que es igual a 0,64. Esto quiere decir que, haciendo 20 comparaciones, tenemos un 64 % de probabilidades de identificar una diferencia como significativa cuando en realidad no lo es o, dicho de otra forma, de cometer un error de tipo I.

Este error aumenta al incrementarse el número de comparaciones, y podemos encontrar una diferencia falsamente significativa solo por azar. Por consiguiente, para la correcta interpretación de los resultados de múltiples pruebas estadísticas es imprescindible aplicar algún método de ajuste del valor de p , como los de Bonferroni, Holm y Simes. De otro modo, para obtener una p significativa, solamente es necesario realizar múltiples pruebas estadísticas. De allí el concepto del filósofo Mario Bunge de que por encima del estándar de oro de la investigación clínica (el estudio randomizado, controlado, doble ciego y estandarizado) está el estándar de platino: las asociaciones y las relaciones tienen que tener sentido biológico.

FALACIA DE P

A partir de la publicación del libro de Ronald Fischer, *Statistical Methods for Research Workers*, en 1925, se popularizó el concepto de significación estadística cuando el valor de p era menor a 0,05. Según parece, esto se debió a una necesidad práctica de la edición del libro, ya que hubiera sido complicado publicar tablas para cada uno de los resultados y Fischer optó por elaborar unas pocas de ellas, entre las que se encontraba la correspondiente a 0,05.

En las décadas de años 40 y 50 del siglo XX las publicaciones de medicina comienzan a utilizar la estadística y el valor de $p < 0,05$ se convirtió en una especie de punto de corte que separó los resultados obtenidos por el simple azar de aquellos que muy probablemente tuvieran importancia y trascendencia fisiopatológica. Pero la significación estadística nada dice de la magnitud del efecto supuestamente detectado: una p significativa puede ocurrir ante una diferencia insignificante o clínicamente irrelevante. Por ejemplo, se puede concluir que un grupo de pacientes sometido a un cierto tratamiento tiene una sobrevida mediana 2 horas mayor que el grupo no tratado. Si bien la diferencia puede ser estadísticamente significativa, resulta evidente que clínicamente carece por completo de importancia. Fue el psicólogo Jacob Cohen quien introdujo la idea de dar relevancia a la magnitud del efecto, del que va a depender, en consecuencia, el tamaño muestral: cuanto menor el efecto buscado, más grande deberá ser la muestra necesaria.

Es evidente que no es lo mismo una diferencia entre medias de determinada variable de 5 o de 500 unidades; y tampoco una asociación entre 2 variables con un r de Pearson de 0,10 que otro de 0,90. La p no nos proporciona esa información. Esta magnitud del efecto debe establecerse antes de realizar el estudio o experimento, lo que influirá en el tamaño muestral necesario y en el poder del estudio. Si el poder del estudio es bajo un resultado negativo, nos puede hacer pensar que la intervención no tuvo ningún efecto cuando en realidad se trata de una especie de «falso negativo» debido o a un tamaño muestral insuficiente o a una diferencia buscada demasiado pequeña.

La relevancia clínica de un fenómeno va más allá de cálculos aritméticos y está determinada por el juicio clínico. La relevancia depende de la magnitud de la diferencia, la gravedad del problema a investigar, la vulnerabilidad, la morbimortalidad generada por el mismo, su costo y por su frecuencia, entre otros elementos.

Otro aspecto a tener en cuenta es que, en teoría, si aumentamos indefinidamente el tamaño muestral podríamos rechazar cualquier hipótesis nula. Es decir que, si la población estudiada es extremadamente grande, aumenta la probabilidad de que la p sea significativa para detectar pequeñas diferencias. Como regla práctica puede decirse que si un estudio se hizo en muchos miles de pacientes (lo que puede parecer *a priori* impresionante) seguramente el

efecto buscado (y encontrado) es pequeño. Nadie reclutaría para un estudio 10 000 pacientes si puede demostrar un efecto con 100.

En ocasiones, por desconocimiento del campo estudiado, no es fácil apreciar en un primer momento qué magnitud del efecto debería buscarse. Resulta interesante en esos casos la llamada «propuesta de Cohen», que es una manera de medir la importancia del efecto o asociación mediante un estadístico denominado diferencia estandarizada de medias o parámetro delta (δ). Este se calcula dividiendo la diferencia entre las medias tipificadas (para el caso de las variables continuas) y el desvío estándar de la variable en cuestión. Por ejemplo, si hay que decidir qué valores de densidad ósea son importantes para evaluar dos grupos diferentes, se calcula el desvío estándar del total de los datos y luego se comparan las medias estandarizadas de ambos grupos. Para el cálculo del tamaño muestral, la magnitud de la diferencia buscada dependerá de cada circunstancia.

Para Cohen, cuando δ es de 0,2 a 0,3 el efecto es pequeño, alrededor de 0,5 el efecto es medio, y lo considera grande cuando es mayor de 0,8.

En resumen, el médico debe estar alerta ante estudios que muestran valores significativos de la p , pero las diferencias son clínicamente irrelevantes.

NNT, NND Y EL RIESGO RELATIVO

Es habitual que la propaganda de la industria farmacéutica, e incluso los trabajos científicos financiados por la misma, reporten que tal o cual droga produce una disminución porcentualmente importante en el riesgo de mortalidad o alguna otra circunstancia indeseable, como infarto de miocardio, ictus, etcétera.

Lo que no se advierte es que se está haciendo énfasis en el riesgo relativo (RR) y no en el riesgo absoluto del evento indeseable.

Un medicamento disminuye en 50 % la mortalidad de una enfermedad (riesgo relativo), pero si el riesgo absoluto es muy bajo (por ejemplo, 1 muerte cada 1000 enfermos) sería necesario tratar a 2000 pacientes para evitar 1 evento. Y si hay 1 efecto adverso fatal cada 500 pacientes, el balance sería claramente negativo: 4 muertos con tratamiento y 1 sin tratamiento.

Por el contrario, si el evento indeseable es muy frecuente en determinada enfermedad y los efectos adversos poco frecuentes, el balance es positivo: en una enfermedad con una mortalidad de 80 % que con cierto tratamiento disminuye al 40 % la disminución del RR sigue siendo del 50 %. Pero ahora solo necesito tratar (NNT) 2,5 pacientes para evitar un evento; y si los efectos adversos fatales ocurren en 1 de cada 1000 pacientes tratados el balance es favorable: casi todos los pacientes tratados se benefician y es muy improbable que alguno se perjudique.

En algunas publicaciones se reportan el NNT y el NND, pero en el caso de que no lo hagan es muy recomendable buscar en el artículo los números absolutos y calcular estos índices.

La mayoría de las drogas que se utiliza en cardiología tiene un NNT de alrededor de 20. Es decir que el médico se enfrenta a un paciente al que le va a prescribir una medicación que va a beneficiar a 1 de cada 20 pacientes que la reciba, y no sabemos cuál va a ser con certeza el beneficiado. Los otros 19 estarán medicados durante años sin beneficio alguno. De esta manera, adquiere relevancia la individualización de los tratamientos y el sopesar con cuidado la relación entre riesgos y beneficios. No es lo mismo tratar un paciente que es solo hipertenso que otro que además presenta otros factores de riesgo vascular. Es probable que en este último el beneficio sea mayor.

Obviamente conocer el NNT cambia la perspectiva del médico sobre la importancia de un tratamiento determinado.

PUNTO FINAL COMBINADO

Cuando la frecuencia de un evento indeseable es pequeña, es común que los estudios utilicen puntos finales (*end point* en inglés) en los que se combinan diversas circunstancias negativas. Por ejemplo, la mortalidad intrahospitalaria del angor inestable es muy baja, por lo que la acción de las intervenciones suele evaluarse mediante el efecto en un punto final que combina muerte, necesidad de cirugía cardíaca, infarto de miocardio y angina refractaria.

Al aumentar la frecuencia del evento combinado (por eso se denomina punto final combinado) es más fácil demostrar un efecto sobre el mismo; recordemos que será necesario un menor tamaño muestral. O, a la inversa, si solamente tuviéramos en cuenta la mortalidad, el tamaño de la muestra debería ser enorme y casi inalcanzable.

El problema es que el resultado puede ser engañoso: una intervención que reduce el punto final combinado podría no tener ningún efecto en el evento de mayor importancia, como, por ejemplo, la mortalidad.

PUNTO FINAL SUSTITUTO

El punto final surrogado (*surrogate endpoint*) o sustituto es aquel evento de interés que los investigadores utilizan cuando el punto final que en verdad interesa es demasiado difícil o costoso de medir, de modo que recurren a otro más fácilmente mensurable que se correlaciona bien con el primero.

Un ejemplo sencillo sería la evaluación del efecto de una medicación para disminuir la presión arterial (PA). En realidad no nos interesa saber qué PA tenemos; lo que nos interesa es mantener nuestra salud cardiovascular.

Cuando la relación entre el punto final sustituto y el evento clínico de interés está suficientemente probada, es razonable utilizar la PA como punto final sustituto; pero no sería lícito que el punto final fuera la elevación de una sustancia en la sangre por un desorden genético si no está probado que su descenso es efectivo para disminuir la morbimortalidad. Un ejemplo sería la enfermedad de Fabry, en la que el déficit congénito de una enzima produce la acumulación en sangre de determinados lípidos que con el tiempo se depositan en diversos órganos y puede llegar a la insuficiencia renal. El tratamiento consiste en el reemplazo por vía endovenosa de la enzima en cuestión sintetizada mediante técnicas de ingeniería genética y un costo de decenas de miles de dólares mensuales durante toda la vida.

Sin embargo, la mayoría de los estudios de eficacia de este tratamiento tiene como punto final la disminución de los niveles en sangre de los lípidos anormales y no la insuficiencia renal o la sobrevida. El estudio que usó un punto final combinado que incluía insuficiencia renal y muerte no fue significativo, y a pesar de ello los autores concluyen en que hubo beneficio luego de realizar un análisis *post-hoc* de un subgrupo de pacientes.

La literatura médica está llena de casos de estudios que utilizan puntos finales sustitutos: la presión ocular para la visión en el glaucoma, el descenso del colesterol para la prevención de eventos vasculares, el intervalo libre de progresión en la mortalidad por cáncer, el número de CD4 y la carga viral en la mortalidad por SIDA, la glucemia en la diabetes en vez de la afectación de órganos blancos, etcétera.

La gran mayoría de las drogas oncológicas dirigidas contra blancos moleculares (anticuerpos monoclonales e inhibidores de tirosinaquinasas) tiene altos costos, pero no ha demostrado ningún aumento de la sobrevida ni mejoría de la calidad de vida y su aprobación por los entes reguladores se ha basado en estudios que evidenciaron mejoría en tiempo a la progresión (punto final sustituto).

En varias oportunidades, en los estudios en los que las intervenciones estaban dirigidas a puntos finales sustitutos, los resultados fueron sorprendentes: la Flecainida hace desaparecer las extrasístoles pero aumenta el porcentaje de muerte súbita (efecto proarrítmico); el Torcetrapib, una droga que aumenta el HDL colesterol, también incrementó la mortalidad en el grupo tratado; la intensificación del control glucémico en la diabetes aumentó la mortalidad; la prevención de determinado tipo de cáncer resultó en mayor mortalidad en el grupo sometido al cribado, etcétera.

Como puede verse, razonamientos fisiopatológicos impecables derivan muchas veces en grandes fracasos al ser llevados al terreno de la comprobación científica. El papel de estas especulaciones biológicas debe circunscribirse

a la generación de hipótesis de trabajo cuya validez debe ser contrastada con los métodos adecuados.

En cada caso, el médico debe tener conciencia de que se está evaluando un punto final que no es el verdaderamente importante y analizar si es apropiado el uso de un punto final sustituto. Esto dependerá de lo que se conozca sobre la enfermedad en cuestión y de la solidez de la evidencia disponible.

FLEXIBILIDAD DE LOS DISEÑOS Y REPORTE SELECTIVO DE RESULTADOS

Es frecuente encontrar combinación de varios diseños, manipulación en el análisis de los datos y maneras de presentar los resultados que conspiran en contra de la calidad de la investigación y de hallazgos que sean confirmados *a posteriori*. Se ha dicho humorísticamente que, si se tortura lo suficiente a los datos, se puede lograr que nos digan lo que queremos.

En ocasiones, los investigadores encuentran resultados positivos utilizando ciertas técnicas y negativos con otras, lo que lleva a reportar solamente los resultados positivos. Esto es, se reportan los resultados en forma sesgada, sin mencionar los resultados negativos.

En el campo de la resonancia magnética funcional cerebral, por ejemplo, al no estar estandarizado el análisis computado de las imágenes obtenidas, se puede encontrar que determinada zona del cerebro se activa en ciertas circunstancias, pero usando otros tipos de análisis se puede afirmar exactamente lo opuesto: que dicha zona no se activa.

Si no se reportan ambos resultados, y por el sesgo que hace más probable la publicación de resultados positivos, la cantidad de estudios con resultados falsos será muy alta; esto ha sido demostrado incluso mediante un análisis matemático por John Ioannidis (1965, profesor de Medicina y Estadística de la Universidad de Stanford).

Cuanto mayor es el número y menos seleccionadas las variables estudiadas, poco probable es arribar a resultados verdaderos. El sesgo de inclusión también es frecuente, se comienza con poblaciones muy numerosas y luego se van seleccionando los pacientes del estudio con distintos criterios, para terminar de formar en la práctica un pequeño subgrupo de la población total. Es conveniente fijarse en el número total de pacientes considerados en un primer momento como candidatos a ingresar al estudio y la cantidad de los que efectivamente lo hicieron. La presencia de grandes diferencias (muchos pacientes excluidos) conspira contra la validez externa del estudio: sus conclusiones poco tienen que ver con los pacientes de la vida real.

Otro aspecto que debe ser analizado con detenimiento en la literatura médica es la correspondencia entre las afirmaciones de las conclusiones y

los resultados realmente obtenidos. A veces los autores llegan hasta a afirmar lo opuesto a lo que se deduce del análisis estadístico.

ANÁLISIS DE SUBGRUPO *POST-HOC*

Los grupos a comparar deben establecerse en el momento del diseño del estudio, antes de realizarlo. Si se divide un gran estudio multicéntrico en 40 subgrupos (por país, edad, sexo, tensión arterial, tratamiento, etc.) tendremos resultados positivos ($p < 0,05$) solamente por el azar en al menos 2 de los subgrupos.

En el estudio ISIS-2 (*Second International Study of Infarct Survival Collaborative Group*), si se hacía un análisis de las fechas de nacimiento de los pacientes en 12 grupos según los signos del zodiaco, se llegaba a la absurda conclusión de que los nacidos bajo los signos de Géminis y Libra no tenían disminución de la mortalidad con el uso de aspirina luego de un infarto de miocardio. El resto de los pacientes, en cambio, mostraba disminución de la mortalidad y un efecto protector de la aspirina.

Otros autores comunicaron un análisis similar en el que todos los pacientes de 6 signos del zodiaco agrupados mostraban una mejor supervivencia a 5 años luego de un trasplante alogénico de médula ósea como tratamiento de la leucemia mieloide crónica. En el análisis multivariado, cuando se ajustó por los factores de riesgo conocidos, la diferencia persistió, con valores de p muy pequeños ($p = 0,007$). Deben desalentarse estos *fishing trips*, o «salir de pesca» que hurgan en los datos con comparaciones no previstas en el diseño inicial.

ESTUDIOS DETENIDOS TEMPRANAMENTE: EFECTOS «INFLADOS»

Cuando un estudio clínico muestra un beneficio que parece muy importante antes de su finalización, se puede considerar que no es ético seguir randomizando pacientes a la rama con placebo. La continuación del estudio sería además un desperdicio de recursos económicos.

Entre la década de 1990 y la de 2000 aumentó a más del doble la cantidad de estudios finalizados tempranamente por efectos beneficiosos en apariencia importantes; esto llevó a que se estudiaran esos ensayos clínicos desde el punto de vista metodológico. Los resultados fueron llamativos, la mayoría de los autores no dio a conocer la manera en que se llegó a la decisión de detener el estudio, como tampoco el tamaño muestral planeado al inicio, la cantidad de análisis interinos realizados o las reglas estadísticas en que se basaron.

Debido al problema de las comparaciones múltiples, si se repiten esos análisis previos (*interim analysis*) la suficiente cantidad de veces, es probable que aparezcan resultados con beneficios grandes por azar y eso lleve a que se detenga el estudio.

Las soluciones a estas cuestiones metodológicas son complejas, por lo que es suficiente que el médico conozca esta posibilidad de error en los estudios que se detienen antes de su finalización.

«VIBRACIÓN DEL EFECTO»: CAMBIO DEL RESULTADO SEGÚN EL ANÁLISIS ESTADÍSTICO

Distintos análisis estadísticos pueden dar resultados contradictorios. Un análisis puede dar resultado negativo si utilizamos un test conservador o resultado positivo si se usa otro test.

Se han comunicado casos extremos en que una prueba determinada, según la manera en que se realiza, puede dar un millón de valores de p que oscilen entre 0,01 y 0,95.

Distintos análisis estadísticos pueden dar resultados contradictorios. Un análisis, dependiendo del test estadístico que se emplea, puede resultar negativo o positivo. Incluso, se han comunicado diversos casos en los cuales para una misma prueba, según la manera en la cual se realiza, los valores de p pueden oscilar entre 0,01 y 0,95.

A pesar de que los softwares estadísticos permiten construir modelos estadísticos predictivos o de simulación, mediante la selección automática de las variables a incluir, es necesario tener presente que esto puede resultar en resultados contradictorios o sin fundamentos teóricos que permitan justificar las observaciones del modelo.

Por lo antes dicho, siempre se debe tener en cuenta la prueba estadística que se ha utilizado para analizar los datos en virtud del efecto biológico que se estudia (plausibilidad biológica).

ESTUDIOS DE NO INFERIORIDAD

En los últimos años ha proliferado este tipo de estudio que no busca demostrar que un tratamiento es superior a otro sino solo que no es peor. Se realizan si se cree que no se puede demostrar la superioridad en un estudio convencional.

Los estudios de no inferioridad tienen sentido desde la lógica del marketing y la competencia de la industria farmacéutica, pero es problemático

justificarlos desde la perspectiva de la ciencia, salvo que el nuevo tratamiento ofrezca ventajas importantes sobre los tratamientos estándar, como menores efectos adversos, menor costo o más practicidad en la dosificación. A menudo no es el caso, y las nuevas terapias tienen costos siderales sin aportar mayores ventajas, tratándose simplemente de una cuestión comercial.

Este tipo de estudio debe analizarse con más cuidado que los de superioridad, ya que errores y problemas pueden sesgar el resultado y hacer más fácil la demostración de no inferioridad. Por ejemplo, seleccionar poblaciones de bajo riesgo que no se benefician demasiado con ningún tratamiento, el no cumplimiento del tratamiento o el cruce de pacientes entre las ramas del estudio y los pacientes perdidos en el seguimiento tienden a que los resultados demuestren más fácilmente la no inferioridad.

SESGO DE PUBLICACIÓN

El primero que advirtió la existencia del mencionado sesgo hace más de 50 años fue Theodore Sterling, quien revisó los artículos publicados en una revista y comprobó que en el 97 % de los escritos que utilizaban pruebas de significación estadística se rechazaba la hipótesis nula, es decir que sus resultados se consideraban positivos y en concordancia con la hipótesis planteada por los autores. Si la casi totalidad de los estudios publicados mostraba resultados positivos, no es lógico pensar que casi ningún investigador haya tenido resultados negativos.

Los estudios exploratorios iniciales sobre cualquier tema suelen ser de un tamaño muestral reducido y por lo tanto es esperable que, por efecto del azar, se produzcan tanto resultados positivos como negativos. Con los modernos métodos de metaanálisis es posible incluso detectar la ausencia de estudios negativos sobre la base de los estudios publicados según la asimetría de su distribución en los gráficos en embudo (*funnel plot*).

Distintos autores han estimado que por cada estudio publicado hay entre 10 y 128 no publicados, y una de las principales razones para que ello ocurra es el sesgo de publicación, ya sea por la actitud de los mismos autores o por el rechazo por parte de los comités editoriales de las revistas científicas. La manera de resolver, al menos en parte, esta cuestión sería llevar un registro de todos los estudios en curso y evitar de ese modo que solo se conozcan los resultados de aquellos con resultados positivos, puesto que es dificultoso identificar en forma retrospectiva la supresión mediante este sesgo de los resultados negativos.

Referencias bibliográficas

- AKOBENG, ANTHONY** (2005). Understanding systematic reviews and meta analysis. *Archives Disease in Childhood*, Aug; 90(8):845–848.
- ALTMAN, DOUGLAS** (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- ÁLVAREZ–MARTÍNEZ, HÉCTOR** (2004). Causalidad en medicina. *Gaceta médica de México*; 140(4):467–72.
- ARCHER, KELLIE** (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. *Computational Statistics & Data Analysis*. May; 51(9):4450–64.
- AUSTIN, PETER; MAMDANI, MUHAMMAD; JUURLINK, DAVID; HUX, JANET** (2006). Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *Journal of Clinical Epidemiology*, Sep; 59(9):964–9.
- AUSTIN, PETER; TU, JACK** (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology*, Nov; 57(11):1138–46.
- BAKER, STUART; KRAMER, BARNETT** (2003). A perfect correlate does not a surrogate make. *BMC Medical Research Methodology*, 3(1):16.
- BASSLER, DIRK; MONTORI, VICTOR; BRIEL, MATTHIAS; GLASZIOU, PAUL; GUYATT, GORDON** (2008). Early stopping of randomized clinical trials for overt efficacy is problematic. *Journal of Clinical Epidemiology*. Mar; 61(3):241–6.
- BATES, DON** (1995). *Scholarly Ways of Knowing: An Introduction in Knowledge and the Scholarly Medical Traditions*. Cambridge: Cambridge University Press.
- BEASLEY, BRENT W.** (2002). Evidence-based Medicine Knowledge, Attitudes, and Skills of Community Faculty. *J Gen Intern Med*. Aug; 17(8):632–640.
- BENDER, RALF; LANGE, STEFAN** (2001). Adjusting for multiple testing – when and how? *Journal of clinical epidemiology*, 54(4):343–9.
- BLAND, MARTIN; ALTMAN, DOUGLAS** (1995). Multiple significance tests: the Bonferroni method. *BMJ*; 310(6973):170.
- BLOOR, DAVID** (1998). *Conocimiento e imaginario social*. Capítulo 1. Barcelona: Gedisa.
- BOARDLEY, DONALD** (1997). Evidence based medicine; a powerful educational tool for clerkship education. *American Journal of Medicine*; 102(5):427–32.
- BOTTASSO, OSCAR** (2013). *Aspectos básicos para la realización de una investigación clínica*. Buenos Aires: Federación Argentina de Cardiología.

- BOURDIEU, PIERRE** (1976). El campo científico. *REDES* 1(2).
- BOUTRON, ISABELLE; DUTTON, SUSAN; RAVAUD, PHILIPPE; ALTMAN, DOUGLAS** (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*; 303(20):2058–64.
- BRADBURN, MICHAEL** (2003). Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. *British Journal of Cancer*; Aug 4;89(3):431–6.
- Survival analysis Part III: multivariate data analysis – choosing a model and assessing its adequacy and fit. *British Journal of Cancer*; Aug 18; 89(4):605–11.
- BUNGE, MARIO** (1978). *La ciencia, su método y filosofía*. Buenos Aires: Ediciones Siglo xx.
- BUTTON, KATHERINE; IOANNIDIS, JOHN; MOKRYSZ, CLAIRE; NOSEK, BRIAN; FLINT, JONATHAN; ROBINSON, EMMA; MUNAFÒ, MARCUS** (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*; Apr 10; 14(5):365–76.
- BUYSE, MARCK; MOLENBERGHS, GEERT; BURZYKOWSKI, TOMASZ; RENARD, DIDIER; GEYS, HELENA** (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*; 1(1):49–67.
- CALLON, MICHEL** (1995). Algunos elementos para una sociología de la traducción: la domesticación de las vieiras y los pescadores de la bahía de St. Brieuc. En Iranzo y col: *Sociología de la ciencia y la Tecnología*. Madrid: csic.
- CANGUILHEM, GEORGES** (1994). *Études d'histoire et de philosophie des sciences: concernant les vivants et la vie*. 5ª ed. Paris: Librairie Philosophique J. Vrin.
- (1999). *Le normal et le pathologique*. Paris: Quadrige PUF.
- CAO, JING; ZHANG, SONG** (2014). Multiple comparison procedures. *JAMA*; 312(5):543–4.
- CARDONA, ÁLVARO** (2008). *Cadáveres, Cementerios y Salud Pública en el Virreinato de Nueva Granada*. Medellín: Universidad de Antioquia.
- CARP, JOSHUA** (2012). On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. *Frontiers in Neuroscience*; Oct 11; 6:149.
- CHALMERS, ALAN** (1990). *La ciencia y cómo se elabora*. Madrid: Siglo XXI.
- CLARK, TAANE G.** (2003). Survival analysis part I: basic concepts and first analyses. *British Journal of Cancer*; Jul 21; 89(2):232–8.
- COHEN, JACOB** (1992). Cosas que he aprendido (hasta ahora). *Annals of Psychology*; 8(1–2):3–18.
- COLLINS, HARRY M.** (1975). The Seven Sexes: A Study in the Sociology of a Phenomenon or the Replication of the Experiments in Physics. *Sociology*, 9(2).

- DRENTH, JOOST** (2008). Case reports: added value counts. *The Netherlands Journal of Medicine*. Jul–Aug; 66(7):289–90.
- EDDY, DAVID** (1990). Practice policies: where do they come from? *JAMA*; 263:1265–75.
- (2005). Evidence-based medicine: a unified approach. *Health Affairs*; 24:9–18.
- ELSTEIN, ARTHUR** (2004). On the origins and development of evidence-based medicine and medical decision making. *Inflammation Research*, 53 Suppl 2:S184–9.
- ESCRIG-SOS, JAVIER** (2005). On how to analyze the credibility of a clinical trial or meta-analysis whose main result is expressed in *odds ratio*, relative risk or hazard ratio. *Cirugía Española*; 78(6):351–6.
- FINCKH, AXEL** (2008). Primer: strengths and weaknesses of meta-analysis. *Nature Clinical Practice Rheumatology*; Mar; 4(3):146–52.
- FOUCAULT, M.** (1994). Crisis de la medicina o crisis de la antimedicina. En *Estrategias de poder*. Barcelona: Paidós.
- GIMENO SACRISTÁN, J. Y PÉREZ GÓMEZ, AI.** (1992). *Comprender y transformar la enseñanza*. Madrid: Ediciones Morata.
- GREEN, BART N.** (2006). How to write a case report for publication. *Journal of Chiropractic Medicine*. Summer; 5(2):72–82.
- GUPTA, SANDEEP** (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*; Jul–Sep; 2(3):109–112.
- GURWITZ, JERRY** (2005). Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ*; Apr 16; 330(7496):895–897.
- HARRELL, FRANK; LEE, KERRY; MARK, DANIEL** (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*; 15:361–87.
- HORTON, LUGANO** (1998). *Histoire de l'intolérance dans le journaux*. *Guggenheim lecture*. Société suisse d'histoire médicale et d'histoire naturelle.
- HUGHES, MICHAEL; POCOCK, STUART** (1988). Stopping rules and estimation problems in clinical trials. *Statistics in medicine*; 7(12):1231–42.
- IOANNIDIS, JOHN; TRIKALINOS, THOMAS** (2005). Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology*; Jun; 58(6):543–9.
- IOANNIDIS, JOHN** (2014). How to Make More Published Research True. *PLoS Medicine*; Oct 21; 11(10):e1001747.
- (2008). Why Most Discovered True Associations Are Inflated: *Epidemiology*; Sep; 19(5):640–8.

- (2005). Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology*; Jun; 58(6):543–9.
- JAESCHKE, ROMAN** (1994). User's guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? *JAMA*; 271:703–7.
- Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA*; 271:389–91.
- KNORR-CETINA, KARIN, D.** (1996). ¿Comunidades científicas o arenas transepistémicas de investigación? Una crítica de los modelos cuasi-económicos de la ciencia. *REDES* 3(7).
- KREIMER, PABLO** (1999). *De probetas, computadoras y ratones*. Capítulo III. Buenos Aires: UNQuilmes.
- KUHN, THOMAS** (1990). *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica.
- LAZCANO-PONCE, EDUARDO** (2001). Case-control epidemiological studies: theoretical bases, variants and applications. *Salud Pública Mex.* Mar-Apr; 43(2):135–50.
- LETELIER, LUZ** (2005). Systematic reviews and metaanalysis: are the best evidence? *Revista Médica de Chile*; 133:246–9.
- LOUIS, PIERRE** (1834). *An essay on clinical instruction. The Nineteenth Century: General Collection: Education*; Pos: Fiche N.1.1.2695.
- MANN, CLIFFORD** (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medical Journal*; Jan; 20(1):54–60.
- MCALISTER, FINLAY A.** (1999). Evidence-Based Medicine and the Practicing Clinician. *Journal of General Internal Medicine*; Apr; 14(4):236–242.
- MERTON, ROBERT** (1942). *La ciencia y la estructura social democrática, en Teoría y estructura social*. México: Fondo de Cultura Económica.
- MOORE, RUCHI; GALVIN, SHELLEY; IMSEIS, HYTHAM** (2006). Thongs, Flip-flops, and Unintended Pregnancy: The Seduction of $P < 0.05$. Available from: http://sys.mahec.net/media/onlinejournal/moore_thongs_flipflops.pdf
- MURILLO, JAVIER** (2006). Retos de la innovación para la investigación educativa. En T. Escudero y A. D. Correa, *Investigación e innovación educativa: algunos ámbitos relevantes*. Madrid: La Muralla, Colección Aula Abierta, 23–54.
- NEUHÄUSER, MARKUS** (2006). How to deal with multiple endpoints in clinical trials. *Fundamental and Clinical Pharmacology*; Dec; 20(6):515–23.

- NUNZIA BETTINSOLI, GIUSE** (2008). Riding the waves of change together: are we all paying attention? *Journal of the Medical Library Association*; Apr; 96(2): 85–87.
- PELLEGRINO, EDMUNDO** (2001). Esencia y sentido de la medicina. En Francesc Torralba i Roselló (ed.) *Filosofía de la Medicina*. Madrid: Institut Borja de bioètica. Fundació Mapfre Medicina.
- PIERSON, DAVID J.** (2009). How to read a case report (or teaching case of the month). *Respiratory Care*; Oct; 54(10):1372–8.
- POLANYI, MICHAEL** (2009). *La lógica de la libertad*. Capítulo 4. Madrid: Katz Editores.
- PRENTICE, ROSS** (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*; 8(4):431–40.
- RISING, KRISTIN; BACCHETTI, PETER; BERO, LISA** (2008). Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Medicine*; 5(11):e217.
- ROSENBERG, WILLIAM** (1995). Evidence-based medicine; an approach to clinical solving. *BMJ*; 310:1122–6.
- SACKETT, DAVID** (1996). Evidence-based medicine; what it is and it aren't. *BMJ*; 312:71–72.
- (2000). *Evidence-based medicine: how to practice and teach EBM*. London. Churchill-Livingstone.
- SAMPIERI, ROBERTO** (2006). *Metodología de la investigación*. 4° Ed. México: McGraw-Hill Interamericana.
- SCHULZ, KENNETH** (2002). Case-control studies: research in reverse. *Lancet*. Feb 2; 359 (9304):431–4.
- STEGERT, MIHAELA; KASENDA, BENJAMIN; VON ELM, ERIK; YOU, JOHN J; BLÜMLE, ANETTE; TOMONAGA, YUKI; SACCILOTTO, RAMON; AMSTUTZ, ALAIN; BENGOUGH, THERESA; BRIEL, MATTHIAS** (2016). An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. *Journal of Clinical Epidemiology*; Jan; 69:152–60.
- STERNE, JONATHAN; SMITH, GEORGE** (2001). Sifting the evidence – what's wrong with significance tests? *Physical Therapy*; 81(8):1464–9.
- STEYERBERG, EWOUT W.; BLEEKER, SACHA E.; MOLL, HENRIËTTE A.; GROBBEE, DIEDERICK E.; MOONS, KAREL G.M.** (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*; May; 56(5):441–7.
- STEYERBERG, EWOUT; BORSBOOM, GERARD; VAN HOUWELINGEN, HANS; EIJKEMANS MARINUS; HABBEMA, DIK** (2004). Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Statistics in medicine*; 23 (16):2567–86.

- STIGLER, STEPHEN** (2009). Fisher and the 5% level. *Chance*, 12; 21(4):12–12.
- TEJEDOR, FRANCISCO** (2008). Innovación Educativa Basada en la Evidencia. I congreso internacional de intercambio de experiencias de innovación docente universitaria. Salamanca, España, Disponible en URL: <https://dialnet.unirioja.es/descarga/articulo/2553098.pdf>
- UMSCHEID, CRAIG** (2011). Key Concepts of Clinical Trials: A Narrative Review. *Postgraduate Medicine*. Sep; 123 (5):194–204.
- YUAN, YUHONG** (2009). Systematic reviews: the good, the bad, and the ugly. *American Journal of Gastroenterology*; May; 104 (5):1086–92.

Sobre los autores

MIGUEL HERNÁN VICCO. Doctor en Medicina (UNR). Magíster en Ciencia, Tecnología y Sociedad (UNQ). Magíster en Docencia Universitaria (UNL). Especialista en Clínica Médica. Médico (UNR). Director del Programa de Desarrollo de la Investigación; Docente del Área de Clínica Médica; Director de la optativa Medicina Basada en la Evidencia en la Práctica Clínica (FCM, UNL). Director de proyectos y miembro de grupos de trabajo. Autor de numerosos artículos en revistas científicas y comunicaciones en congresos.

LUZ MARÍA RODELES ANTONELLI. Médica (UNR). Especialista en Clínica Médica. Doctoranda en Ciencias Biomédicas (UNR). Becaria doctoral del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Docente del Área Clínica Médica (FCM, UNL). Integrante de proyectos de investigación, codirectora de pasantías, becas de iniciación a la investigación y Proyectos de Extensión de Interés Social. Coordinadora de Proyectos de Investigación (FCM, UNL). Autora de numerosos artículos en revistas científicas y comunicaciones en congresos.

LARISA IVÓN CARRERA. Médica (Facultad de Ciencias Médicas, UNR). Especialista en Dermatología. Especialista en Medicina Legal (UNL). Magister en Educación Universitaria (UNR). Doctora (UBA). Prof. Titular Biología Celular, Histología y Embriología (FCM, UNL). Decana (FCM, UNL). Directora de diversos proyectos de Extensión. Representante de la UNL Comité Académico de Atención Primaria de la Salud (AUGM). Miembro de la Comisión Académica de la carrera de especialización en Vinculación Tecnológica (UNL).

HÉCTOR MARIO MUSACCHIO. Médico. Especialista en Clínica Médica. Médico Especialista en Oncología. Doctor en Medicina. Jefe del Servicio de Clínica Médica del Hospital Iturraspe. Profesor Adjunto de Clínica Médica y Terapéutica I y II (Unidad Docente Santa Fe dependiente de la Facultad de Ciencias Médicas, Universidad Nacional de Rosario). Profesor Titular de Medicina Interna (Facultad de Ciencias Médicas, Universidad Nacional del Litoral).

MELISA LUCIANA PONCE. Estudiante de Licenciatura en Diseño de la comunicación Visual, orientación Diseño Gráfico (Facultad de Arquitectura, Diseño y Urbanismo, UNL). Miembro del Área Diseño y Audiovisual-Ceremonial (Facultad de Ciencias Médicas, UNL). Colaboradora desde el área de Diseño del Programa «Equidad en Salud» (FCM, UNL).